

661 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
 b 2197 GCGAGGCAAGACTGCTGCGCCGGTCAAGAGCGTGCAGATCATCCGCACTTCTGCT 2256  
 y 681: ProLeuLysGlnTyrPheAlaCyVal 689  
 b 2257 CCGCTGAAGAGATATTTGCGTGTGTG 2283

RESULT 2  
 S-10-144-577-2  
 Sequence 2, Application US/10144577  
 Publication No. US20030083292A1  
 GENERAL INFORMATION:

APPLICANT: Macleod, Alan Robert  
 TITLE OF INVENTION: Inhibitors of DNA Methyltransferase Isoforms  
 FILE REFERENCE: MET-005  
 CURRENT APPLICATION NUMBER: US/10/144,577  
 PRIOR FILING DATE: 2002-05-13  
 PRIOR APPLICATION NUMBER: US 60/290,202  
 PRIOR FILING DATE: 2001-05-14  
 PRIOR APPLICATION NUMBER: US 60/290,212  
 PRIOR FILING DATE: 2001-05-11  
 NUMBER OF SEQ ID NOS: 49  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 3005  
 TYPE: DNA

ORGANISM: Homo sapiens  
 S-10-144-577-2

Alignment Scores:  
 Ref. No.: 0 Length: 3005  
 Score: 3763.00 Matches: 689  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Very Match: 100.0% Indels: 0  
 Gaps: 0

S-10-623-813-86 (1-689) x US-10-144-577-2 (1-3005)

1 MetAenAlaValIGluGlnGlnGlyProGlyGlnSerGlnLysValGluGlnAlaSer 20  
 899 ATGAATGCTGTGGAAGAAACCAAGGGGCGCGGAGTCTCAAGAGTGGAGAGGCGCAC 958  
 21 ProProAlaValIGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40  
 959 CCTCTGCTGTGCACAGCCCACTGACCCCGCATCCCACTGTGCTACACGCTGTGAG 1018  
 41 ProValIGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
 1019 CCCGTGGGTCCGATGCTGGGAGCAAGATGCCACCAAGCAGCGGATGACGAGCCAGAG 1078  
 61 TyrGluAspGlyArgGlyPheGlyTyrIleGlyGluLeuValThrGlyLysLeuArgGlyPhe 80  
 1079 TACGAGGAGCGCGGGGCTTGTGGCAATGGGAGCTGTGTGGGGAATCTGGGGGCTTC 1138  
 81 SerTrpTrpProGlyArgGlyLeuValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
 1139 TCTCTGTGGCCAGGCGCGCATGTGTCTTGTGTGATACCGGCGGAGCCGAGCAGCTGAA 1198  
 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCyValGluLys 120  
 1199 GGCACCCGCTGGGTCAATGTGTTCGAGACCGCAATTCATCATGTGTGTGTGTGAGAG 1258  
 121 LeuMetProLeuSerSerPheCySerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140  
 1259 CTGATCCCGCTGAGCTCTTTTGCAGTGGCTTCACACGAGCCACGTACAAACAGAGCC 1318  
 141 MetTyrArgLysAlaIleTyrGlyValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 1319 ATGTACCGCAAGCCATCTACAGAGTCTTCAGGTGCGCAGCAGCCCGCGGGGAGAGCTG 1378

Qy 161 PheProValCyHisIleAspSerAspGluSerAspThrAlaLysAlaValGluGlnAsn 180  
 Db 1379 TTTCCGCTGTGCCACACAGCCATGAGAGTACATCCCAAGCCCTGTGAGGTGCAGAAC 1438  
 Qy 181 LysProMetIleGlnTyrPheAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
 Db 1439 AACCCATGATTAATGTGGCCCTCGGGGGCTTCACGCTTCTGTGGCTTAAGGGCTGTGAG 1498  
 Qy 201 ProProGluGlnGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro 220  
 Db 1499 CCACAGAGAAGAGAAGATCCCTCAAAAGAGTGAACGAGCATGTGGTGGAACT 1558  
 Qy 221 GluAlaAlaAlaTyrLysLysProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
 Db 1559 GAGGAGCTGCTCAACCACTCCACCAAGCCCAAGCCCGGAGAGACACAGCCGAG 1618  
 Qy 241 LysProLysValLysGlnIleIleAspGluArgThrArgGlnArgLeuValTyrGluVal 260  
 Db 1619 AACCCAAAGTCAGAGAGATTTATGTAGCCGACCAAGAGCGGCTGTGTACAGAGTG 1678  
 Qy 261 ArgGlnLysCyArgAsnIleGluAspIleCyHisSerCyGlySerLeuAsnValThr 280  
 Db 1679 CCGCAGAGTGCAGGAAATTTGAGCATCTGCATCTCTGTGGAGCCCAATGTACC 1738  
 Qy 281 LeuGlnHisProLeuPheValGlyGlyMetCyGlnAsnCyLysAsnCyPheLeuGlu 300  
 Db 1739 CTGGAACACCCCTCTTCTGTGAGGAATGTGCCAAACCTGCAGAACTGCTTCTGAG 1798  
 Qy 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCySerThrIleCyGlyGly 320  
 Db 1799 TGTGCTTACCAATGACACACAGCGCTACAGTCTTACTGACCACTGTGTGTGGGGG 1858  
 Qy 321 ArgGluValLeuMetCyGlyLysAsnAsnAsnCyGlyArgCyPheCyValGluCyVal 340  
 Db 1859 CGTGAAGTCTCATGTGGGAAACAACAACCTCTGAGGTGCTTTTGGGTGAGTGTGTG 1918  
 Qy 341 AspLeuLeuValGlyProGlyValAlaGlnAlaIleLysGluAspProTrpAsnCy 360  
 Db 1919 GACCTTGTGTGGGCGCGGGGCTGCCGAGCAGCATTAAGAGAACCCCTGGAATCTCG 1978  
 Qy 361 TyrMetCyGlyHisLysGlyTyrThrGlyLeuLeuArgArgArgGlnAspTrpProSer 380  
 Db 1979 TACATGTGGGACACAGAGGATCTTACGGGCTGTGCGCGGCGAGAGGACTGGCCCTTC 2038  
 Qy 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400  
 Db 2039 CGGCTCCAGATTTTGTGCTTAATAACACAGCAGAAATTGACCTTCAAGGTTTAC 2098  
 Qy 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
 Db 2099 CCACTGTCCACAGCTGAGAGAGAGAACCCATCCGGGTGTGTCTCTTGTATGGAATC 2158  
 Qy 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyTyrIleGlnValAspArgTyrIleAlaSer 440  
 Db 2159 GCTACAGGCTCTCGTGTCTGAGAGCATTTGGGCAATTCAGGTGACCGCTACATTCCTG 2218  
 Qy 441 GluValCyGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
 Db 2219 GAGGTGTGAGAGTCCATTCACGCTGGCATGTGTGCGCACAGGGGAGAGATCATGTAC 2278  
 Qy 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeu 480  
 Db 2279 GTGGGAGAGTCCGACGCTCACAGAGATATTCAGAGAGTGGGCCCATTCGATCTG 2338  
 Qy 481 ValIleGlyLysSerProCyAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db 2339 GTGATTGGGGGAGTCCCTGATGATGACCTTCATCTCAACCTGCTGCAAGGGGCTC 2398  
 Qy 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
 Db 2399 TACAGAGGACAGCGGCTCTTCTTGTGAGTTCTACCGCTCTGTGATATGTGGCGGCC 2458  
 Qy 521 LysGluGlyLysAspAspArgProPhePheThrLeuPheGluAsnValAlaMetGlyVal 540

us-10-623-813-8

RESULT 3  
US-10-172-118-132  
Sequence 132, Application US/10172118

Db	2039	CGGCTCCAGATGTTCTTCGCTAATAACACAGACCAGAAATTTGACCTCCAAAGTTTAC	2098
Qy	401	ProProValProAlaGluysArgLysProIleArgValLeuSerLeuPheAspGlyIle	420
Db	2099	CCACCTGTCCTCCAGCTGAGAAAGAGAGGCCATCCGGGTGCTGTCCTTTGATGGAAATC	2158
Qy	421	AlaIrnGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer	440
Db	2159	GCTACAGGGCTCCTGTGTGCTGAAGAGCACTTGGGCAATTCAGGTGAGCCGCTACATTTGCTCG	2218
Qy	441	GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr	460
Db	2219	GAGGTGTGTGAGGACTTCATCAGGTGGGCAATGTGTGGCGGACCAAGGGAAATCATGTATC	2278
Qy	461	ValGluAspValArgSerValIleThrGlnLysHisIleGlnGlnTyrPheProPheAspLeu	480
Db	2279	GTGGGGAGAGCTCCGCAAGCTCACACAAAGCATATCCAGAGTGGGGCCCATTTCCATCTCG	2338
Qy	481	ValIleGlyGlySerProCysAsnAspLeuSerIleValLeuProAlaArgLysGlyLeu	500
Db	2339	GTAATTGGGGGAGTCCCTGCAATGACCTTCATCTGCAACCTGTGCGGACAGGGGCTTC	2398
Qy	501	TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro	520
Db	2399	TACGAGGGCACTGGCCGGCTCTTCTTAAGTTCTACCCGCTCTTCATGATGTCGGGGCCC	2458
Qy	521	LysGluGlyAspAspArgProPhePheThrLeuPheGluAsnValValAlaMetGlyVal	540
Db	2459	AAGGAGGAGATGATTCGCCCTTCTTCCTTGCGCTTGTGAAGATGTGTGGCCATGGGCGTT	2518
Qy	541	SerAspLysArgAspIleSerArgPheLeuGlnSerAsnProValMetIleAspAlaLys	560
Db	2519	AGTGACAAAGAGGAGCATCTCGGATTTCTCGAGTCCAAACCTGTGATGATGATGCGCAA	2578
Qy	561	GluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg	580
Db	2579	GAAAGTGTAGCTGCACACAGGGCCCGCTACTTCTGGGGTAACTTCCCGGTATGAACAG	2638
Qy	581	ProLeuAlaSerThrValAsnAspLysLeuGlnGluCysAlenGlnHisIleGlyArg	600
Db	2639	CCGTTGGCATCCACTGTGAATGATTAAGCTGAGAGCTCAGAGAGTGTGGAGCATGGCAGG	2698
Qy	601	IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly	620
Db	2699	ATAGCCAAAGTTACGAAAGCTGAGGACCATTACTACAGAGCTCAACCTCATTAAAGCAGGGC	2758
Qy	621	LysAspGlnHisPheProValPheMetAsnGlnLysGluAspIleLeuTyrCysThrGlu	640
Db	2759	AAAGACACGACATTTTCTGTCTTCAATGATGAAAGAGAGACATCTTAATGTGACATGAA	2818
Qy	641	MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu	660
Db	2819	ATGAAAGGGAATTTGGTTTCCAGTCCACTAATCACTACCTCTCCCAACATGAGCGGCTTG	2878
Qy	661	AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla	680
Db	2879	GGAGGCAAGACTGTGTGGGCGGGTCATGAGCGTGCAGTCAATCCGACACTTTCGCT	2938
Qy	681	ProLeuLysGluTyrPheAlaCysVal	689
Db	2939	CCGCTAAAGAGATTTTTCGCTGTGTG	2965

REFERENCE AUTHORS	TITLE	JOURNAL PUBMED	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominae; Homo. 1 (baaes 1 to 3005) Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K and Li,E.	Cloning, expression and chromosome locations of the human DNMT3 gene family	Gene 236 (1), 87-95 (1999) 2 (baaes 1 to 3005) 10433969	Xie,S., Okano,M. and Li,E. Direct Submission Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th Street, Charlestown, MA 02129, USA 3 (baaes 1 to 3005) Hata,K., Shirohzu,H., Sasaki,H. and En,L. Direct Submission Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th Street, Charlestown, MA 02129, USA	Sequence update by submitter On Feb 12, 2001 this sequence version replaced gl:4927369. 1 3005 Location/Qualifiers

```

source
1:..3002
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p23; between D2S171 and D2S174; 48-50 cM"
1:..3005
gene
/feature="DNMT3A"
1:..3058
CDS
230..2958
"DNMT3A"

```

```

//gene="Umn13a
//function="de novo DNA cytosine methyltransferase"
/codon_start=1
/product="DNA cytosine methyltransferase 3 alpha"
/protein_id="AAD31084.2"
/db_xref="GI:12746532"
/translation="MPAMPSSGGPDITSSNAAREEDRKQGEDEEPGRKEEPEPSTT
ARKGPRGRKRKHPVEGSDTPKDPDAIVSISPSMAODSGASSELLPNDDLEKRSBPQE
EGSRAGQKGTAPAEAGGAALTEPLAERAVENCCCTTEERGGAPAEAGKEQKTNIES
KMGESKSRRLRGSGI GSGESSLRPMPRLTEQAGPVTIYSKRDEDMILAMKREAEKA
KVIAGMNAVENBQGPRESOKVEAPPAVQODTPDPAFTVAETPEAGSAGCKNAIK
AGDEPEYEDERGGFTIGELVNGKIRGSSVMPMPGRIVSMWMTGRSAAGTVMWVFCGG
KESVVCPEKMLPFCFSAPHQATYNKQPMTRKALIVYLIVASRAKCLPVCSDSDPE
SDTAAAEVQNKPMLEWALGGFOPSGEGALVEPEEKQPIKEYATITDMMWPEAAVAAR
PSTAPKPKSTAEAKPKVAILIDERTKRLIVELYAKCKRNIIDITSCGSLNVTITLHEHL
FVGGCQCNKCNKFLCAAYQDDYOSCTICCGGAEVLMGANNCCRCFEVECDVILL
VQPGAQAAIKEDPNNCKMGHKGTGYLLRRREDDPRLQMFANNNDQDEDPKVTILL
PVPKAKRPIRIVLSFDGATIGVLSKDIQIVRYIASEVCSITIVGMVRRHQGKITM
YVGDRASTTQGHIGDWBQEPDILVIGGSPNDQSIYNAPARKGLYEGTGLIFEFYRLIHD
PEMNNPILASTVNDKLEOELCHGRIALRQVLRGSRMSVPYIRRLPAPLKEYFACV"
DILMCTEMERVFGEFVATVDSNNRSLRKLQKVLGRGSRMSVPYIRRLPAPLKEYFACV"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 0
Score: 3763.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 100.0%
DB: 5
Gaps: 0
US-10-623-813-86 (1-689) x AF067972 (1-3005)
Oy 1 MetAsnAlaValGluGlnAsnGlnGlyProGlyGluSerGlnuValGluGluAlaSer 20
Db 899 ATGAATGCTGTGGAGAAACCCAGGGGCCCGGGGAGTCTCAGAGAGTGGAGAGCCAGC 958
Oy 21 ProProAlaValGluGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40
Db 959 CCTCTGCTGTGTGACGACGCCCACTGACCCCGCATCCCCACATGCTGTGCTACCAAGCCTGAG 1018

```

Db	Accession	Gene	Protein	Length (aa)
Db	2059	CCACTGTCCTCCAGCTGAGAAAGGAGGCCATCCGGGAGCTGTCTCTCTTGAATGGAATC		2158
Qy	421	AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAsnSer		440
Db	2159	GCTACAGGGCTCTCTGATGCTGAAAGACATTGGGCATTCAAGTGAACCGGTACATTGGCTCG		2218
Qy	441	GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr		460
Db	2219	GAGGTGTGTGAAGATCTTCATCAACGGTGGGCATGATGGCCGACACAGGGGAGATCATGTAC		2278
Qy	461	ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeu		480
Db	2279	GTCGGGGAAGCTCCGACGCTCCACAGAGCAATATCCAGAGTGGGGCCCATTCATCTCG		2338
Qy	481	ValIleIleGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu		500
Db	2339	GTGATTGGGGGAGTGCTCTGCAATGACCTCTTCATGTGTAACCTGTCTGCAAGGGCCCTC		2398
Qy	501	TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro		520
Db	2399	TACGAGGCACTGGCCGGCTCTCTTCTTGAATCTACCGGCTCCGTCATGATGGCCGGCC		2458
Qy	521	LysGluGlyAspAspArgProPhePheTyrPheGluAsnValValAlaMetGlyVal		540
Db	2459	AAGAGAGGAGATGATATGCCCCCTCTTCTGTGCTTTGAGAATGTGGTGGCCATGGGCGTT		2518
Qy	541	SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys		560
Db	2519	AGTACAGAGGGACATCTCGCATTTCTTCGAGTCCACCTGTGATGATTGATGCCAA		2578
Qy	561	GluValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArg		580
Db	2579	GAAGTGCAGCTGCACACAGGGCCCGTACTTCTGGGGTAACTTCCCGGTATGACAGC		2638
Qy	581	ProLeuAlaSerThrValAsnAspLysLeuGlnGluCysLeuGluHisIleGlyArg		600
Db	2639	CCGTGGCATCCACTGTGATGATGATTAAGCTGGAGCTGAGAGGTCTGAGAGCATGGCAGC		2698
Qy	601	IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly		620
Db	2699	ATAGCCAAAGTTCAGCAAGTGAAGACCATTAACAGAGCTCAAACTCCATTAAGCAAGGC		2758
Qy	621	LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGlu		640
Db	2759	AAAGACCAAGCATTTTCTCTGCTTCATGAATGAGAAGAGACATCTTAATGTGCACGAA		2818
Qy	641	MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu		660
Db	2819	ATGAAAAGGGTATTTGGTTTCCAGTCCACTTAATGACGCTTCCAACTAAGAGCCGCTTG		2878
Qy	661	AlaArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAla		680
Db	2879	GCGAGGAGAGACTGCTGGCCGGTCAATGAGCTGCGCATGCAATCCGCACCTTTCGCT		2938
Qy	681	ProLeuLysGlyTyrPheAlaCysVal		689
Db	2939	CCGCTGAAGAGTATTTTGGCTGTGTG 2965		
RESULT 4	CS050242	3604 bp	DNA	linear
LOCUS	CS050242			
DEFINITION	Sequence 26 from Patent WO2005021757.			
ACCESSION	CS050242			
VERSION	CS050242.1	GI:61889497		
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			
REFERENCE	1			
AUTHORS	Tomme, P.H. and van Rompaey, L.			

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OW protein - nucleic search, using frame\_p2n model

Run on: October 5, 2006, 18:12:34 ; Search time 10519 Seconds  
(without alignments)  
6282.884 Million cell updates/sec

Title: US-10-623-813-86

Perfect score: 3763  
Sequence: 1 MNAAVENOGSGESQKVEAS.....WSVPVIRHLFAPLKEYFACV 689

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 6366136 segs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame-p2n.model -DRV=xlh  
-Q=/abs/ABSSWEB.spool/US10623813/runat\_05102006\_121146\_27659/app\_query.fasta\_1  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdd -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEADSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss05h  
-USER=US10623813 @CCN\_1\_1\_6706 @runat\_05102006\_121146\_27659 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -XGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

GenEmbl: \*  
1: gb\_env: \*  
2: gb\_pat: \*  
3: gb\_ph: \*  
4: gb\_pl: \*  
5: gb\_pr: \*  
6: gb\_ro: \*  
7: gb\_sts: \*  
8: gb\_sy: \*  
9: gb\_un: \*  
10: gb\_vi: \*  
11: gb\_ov: \*  
12: gb\_hcg: \*  
13: gb\_in: \*  
14: gb\_cm: \*  
15: gb\_ba: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3763	100.0	2371	5	AF480163 Homo sapi
2	3763	100.0	3005	2	CS050243 Sequence
3	3763	100.0	3005	5	AF067972 Homo sapi

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	AUTHORS	JOURNAL	PUBMED	REFERENCE	TITLE	AUTHORS	JOURNAL	FEATURES
AF480163	AF480163	Homo sapiens DNA cytosine methyltransferase 3A2 (DNMT3A2) mRNA, complete cds.	AF480163	AF480163	AF480163.1 GI:23954437		Homo sapiens (human)										
CS050242	CS050242	Sequence	CS050242	CS050242	CS050242.1		Homo sapiens (human)										
AF331856	AF331856	Homo sapi	AF331856	AF331856	AF331856.1		Homo sapiens (human)										
DD171352	DD171352	Marker fo	DD171352	DD171352	DD171352.1		Homo sapiens (human)										
BC043617	BC043617	Homo sapi	BC043617	BC043617	BC043617.1		Homo sapiens (human)										
CS050241	CS050241	Sequence	CS050241	CS050241	CS050241.1		Homo sapiens (human)										
CO720469	CO720469	Sequence	CO720469	CO720469	CO720469.1		Homo sapiens (human)										
AF480164	AF480164	Mus muscu	AF480164	AF480164	AF480164.1		Mus musculus (house mouse)										
BC007466	BC007466	Mus muscu	BC007466	BC007466	BC007466.1		Mus musculus (house mouse)										
AF068625	AF068625	Mus muscu	AF068625	AF068625	AF068625.1		Mus musculus (house mouse)										
AY271299	AY271299	Bos tauru	AY271299	AY271299	AY271299.1		Bos taurus (domestic cattle)										
AB208833	AB208833	Homo sapi	AB208833	AB208833	AB208833.1		Homo sapiens (human)										
AB214886	AB214886	Gallus ga	AB214886	AB214886	AB214886.1		Gallus gallus (chicken)										
AC120824	AC120824	Rattus no	AC120824	AC120824	AC120824.1		Rattus norvegicus (brown rat)										
AC095441	AC095441	Rattus no	AC095441	AC095441	AC095441.1		Rattus norvegicus (brown rat)										
AB196917	AB196917	Danio rer	AB196917	AB196917	AB196917.1		Danio rerio (zebrafish)										
AC112586	AC112586	Rattus no	AC112586	AC112586	AC112586.1		Rattus norvegicus (brown rat)										
AB196919	AB196919	Danio rer	AB196919	AB196919	AB196919.1		Danio rerio (zebrafish)										
AC112040	AC112040	Rattus no	AC112040	AC112040	AC112040.1		Rattus norvegicus (brown rat)										
AK025230	AK025230	Homo sapi	AK025230	AK025230	AK025230.1		Homo sapiens (human)										
AB196915	AB196915	Danio rer	AB196915	AB196915	AB196915.1		Danio rerio (zebrafish)										
AC009474	AC009474	Homo sapi	AC009474	AC009474	AC009474.1		Homo sapiens (human)										
AF068627	AF068627	Mus muscu	AF068627	AF068627	AF068627.1		Mus musculus (house mouse)										
AF151974	AF151974	Mus muscu	AF151974	AF151974	AF151974.1		Mus musculus (house mouse)										
AF151970	AF151970	Mus muscu	AF151970	AF151970	AF151970.1		Mus musculus (house mouse)										
AF068626	AF068626	Mus muscu	AF068626	AF068626	AF068626.1		Mus musculus (house mouse)										
AF151973	AF151973	Mus muscu	AF151973	AF151973	AF151973.1		Mus musculus (house mouse)										
AF151969	AF151969	Mus muscu	AF151969	AF151969	AF151969.1		Mus musculus (house mouse)										
AF176228	AF176228	Homo sapi	AF176228	AF176228	AF176228.1		Homo sapiens (human)										
BC105922	BC105922	Mus muscu	BC105922	BC105922	BC105922.1		Mus musculus (house mouse)										
CO726303	CO726303	Sequence	CO726303	CO726303	CO726303.1												
AF156488	AF156488	Homo sapi	AF156488	AF156488	AF156488.1		Homo sapiens (human)										
CQ981515	CQ981515	Sequence	CQ981515	CQ981515	CQ981515.1												
AF331857	AF331857	Homo sapi	AF331857	AF331857	AF331857.1		Homo sapiens (human)										
DO321787	DO321787	Homo sapi	DO321787	DO321787	DO321787.1		Homo sapiens (human)										
AY244710	AY244710	Bos tauru	AY244710	AY244710	AY244710.1		Bos taurus (domestic cattle)										
AB196918	AB196918	Danio rer	AB196918	AB196918	AB196918.1		Danio rerio (zebrafish)										
AY244712	AY244712	Bos tauru	AY244712	AY244712	AY244712.1		Bos taurus (domestic cattle)										
AF068628	AF068628	Mus muscu	AF068628	AF068628	AF068628.1		Mus musculus (house mouse)										
AF151976	AF151976	Mus muscu	AF151976	AF151976	AF151976.1		Mus musculus (house mouse)										
AF151971	AF151971	Mus muscu	AF151971	AF151971	AF151971.1		Mus musculus (house mouse)										
AY078427	AY078427	Mus muscu	AY078427	AY078427	AY078427.1		Mus musculus (house mouse)										

#### ALIGNMENTS

RESULT 1  
AF480163  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
TITLES  
AUTHORS  
JOURNAL  
PUBMED  
REFERENCE  
TITLES  
AUTHORS  
JOURNAL  
FEATURES

AF480163  
Homo sapiens DNA cytosine methyltransferase 3A2 (DNMT3A2) mRNA, complete cds.  
AF480163  
AF480163.1 GI:23954437  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2371)  
A Novel Dmrt3a Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Active de Novo Methylation  
J. Biol. Chem. 277 (41), 38746-38754 (2002)  
2 (bases 1 to 2371)  
Chen, T. and Li, E.  
Direct Submission  
Submitted (31-JAN-2002) Cardiovascular Research Center, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA  
Location/Qualifiers

```

source
1. .2371
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p23"
1. .2371
/gene="DNMT3A2"
217. .2286
/function="de novo DNA cytosine methylation"
/codon_start=1
/product="DNA cytosine methyltransferase 3A2"
/protein_id="AA04037.1"
/db_xref="GI:23954438"
/translation="MNAVEENQGESQKVEEAPPAVOQPTDPASPVTATTPPEVGS
DAGKNATKADDEPEYEDGRGFGIGELVWGLKRFMSWPKRIYSMMWTGSRABET
RWVMFQDGKFSVVCVEKLPMSFCSAFHQATYKQPMYKAIYEVLQVASSRAGKL
FPVCHDSDESDTAKAEVONKPMIEMALGFQPSGPKLEPPEEKKNYEKVTIDMV
EPBAAYAPPPAPKPKPKRSTAEKPKVKIIDEIRTERLVYEVROKCRNIEDICSGS
LNTLLEHPLFYGKGNCKNCFLECAVYDDQSYCTICCGREVLMCGNNCCRC
FCVCCVDLVGPAAQAIKEDPMVCWCGKGYGLLRREDSRLQMFANNHDO
EFDPKXYPPVPAPKRPRIYVLSLFDGIATGLVKDGIQVDRIASEVEDS1TVG
MVRHQKIMYGVDSVYQKHIQEWGPDLYGSPQNDLSIVNPARKGLYEGTGRLF
FEFVRLHDARKEGDDRPFMLPENVMAMGVSIDKDISREFLESNPMVIDAKESAAH
PARYFMKNLPGMNRPLASTVNDKLEQCELEHRIAKRSKRITITTSNSIKQKDOH
PVPVMEKEDILNCTEMERVRFPVHYTIDVSNMSRLARQLLGRSMSPVIRLHAPL
KEYPACV"

Alignment Scores:
Score: 0 Length: 2371
Percent Similarity: 3763.00 Matches: 689
Best Local Similarity: 100.0% Conservative: 0
Query Match: 100.0% Mismatches: 0
Indels: 0
Gaps: 0

US-10-623-813-86 (1-689) x AF480163 (1-2371)
QY 1 MetAsna1aValGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGlnAlaSer 20
Db 217 ATGATGCTGTGGAAGAAACCGAGGGCCCGGGAGCTCTCAGAAAGCTGAGAGGCCAGC 276
QY 21 ProProAlaValGlnGlnInProThrAspProAlaSerProThrAlaAlaThrThProGlu 40
Db 277 CCTCTGCTGTGAGCAGCAGCCACCTGACCTCCCTGCTGCTACCAAGCTGAG 336
QY 41 ProValGlySerAspAlaGlyAspLysAsna1aThrLysAlaGlyAspAspGluProGlu 60
Db 337 CCCGTGGGCTCGATGCTGGGGACAGAAATGCCACCAAGCAGGAGAGAGCCAGAG 396
QY 61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPhe 80
Db 397 TACGAGAGCGGCGGGGCTTGGCAATGGGAGCTGTGGGGGAAACTCGGGGGCTTC 456
QY 81 SerTrrPrrProGlyArgIleValSerTrrPrrMetThrGlyArgSerArgAlaAlaGlu 100
Db 457 TCCGTGTGGCCAGGCGGCAATGTGTCTTGTGTGATGAGCGGGCCGAGCGAGAGCTGAA 516
QY 101 GlyThrArgTrrPrrAlaMetTrrPheGlyAspGlyLysPheSerValCysValGluLys 120
Db 517 GGCACCCGCTGGGTCAATGCTTGGGAGACCGCAAAATCTCAGTGTCTGTGTGAGAG 576
QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140
Db 577 CTGATGCCGCTGAGACTCGTTTGGCAGTGTCTCCACGAGCCACGTACCAACAGCAGCCC 636
QY 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160
Db 637 ATGTAACGCAAGCAATCTACGAGGTCTTGACAGGTGCGCACAGCGCGGCGGAGAGCTG 696
QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluAlaAsn 180

```

```

Db 697 TTCCGGGTGTGCCACGACGAGATGAGATGACCTGCCAAGCGCGTGGAGGTCCAGAAC 756
QY 181 LysProMetIleGluTrrPrrAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200
Db 757 AAGCCATGATTGAATGGGCGCTTGGGGGCTTCCAGCCCTTGTGCGCTTAAGCGCTGGAG 816
QY 201 ProProGluGluGluLysAsnProTrrLysGluValTyrThrAspMetPrrValGluPro 220
Db 817 CCACAGAGAGAGAGAAATCCCTTACAAAGATGTAACGACATGTGGTGGAACTT 876
QY 221 GluAlaAlaAlaTrrAlaProProProAlaLysLysProArgLysSerThrAlaGlu 240
Db 877 GAGCAGCTGCTGTACGACACCTCCACGACCAAAAAGCCCCGAGAGAGCAGCGG1.3 936
QY 241 LysProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluVal 260
Db 937 AAGCCAAAGTCAGAGGATTAATGATGAGCGCAAGAGAGCGGCTGTGTACGAGGT 996
QY 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValTrr 280
Db 997 CGGCAAGTGTCCGGAACATTGAGACATCTGCATCTCTGTGGAGCTCAATGTTACC 1056
QY 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300
Db 1057 CTGGAACACCCCTCTTGTGAGGAAATGTGCCAAACTGCAAGAACTGCTTCTGAG 1116
QY 301 CysAlaTyrGlnTrrAspAspArgLysTrrGlnSerTrrCysThrIleCysCysGlyGly 320
Db 1117 TGTGCTGACAGATACGACGACGAGGCTACGACGCTCACTGCAACATCTGCTGTGGGGC 1176
QY 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340
Db 1177 GGTAGGTGCTCATGTGTGGAAACACATCTGCTGAGGTCTTTTGTGGTGAATGTGTG 1236
QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTrrAsnCys 360
Db 1237 GACCTTGTGGGGGCGGGGGCTGCGCAGCAGCATTAAAGAAACCCCTGGAATGCG 1296
QY 361 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgArgGluAspTrrProSer 380
Db 1297 TACATGTGCGGCAACAGGATACCTGAGGCTGTGGCGGCGGACAGAGACTGGCCCTCC 1356
QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLysValTyr 400
Db 1357 CGGCTTCAGATGTTCTTCTGCTAATACCAAGACCAAGAAATTGACCTCCAAAGTTTAC 1416
QY 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420
Db 1417 CCACCTGTCCACGCTGAGAAAGAGAACCCATCCGGGTGCTGTCTCTTTGATGAGATC 1476
QY 421 AlaThrGlyLeuLeuValIleLysAspLeuGlyIleGlnValAspArgTrrIleAlaSer 440
Db 1477 GCTACAGGGCTCTCGTGTGTGAAGACTTGAGCAATTGAGTGAGCCCTCATATTGCCCTCG 1536
QY 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460
Db 1537 GAGGTGTGTAGACTTCATCACGGTGGCATGTGGGACACCAAGGAGATGATGTATAC 1596
QY 461 ValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 480
Db 1597 GTCGGGAGACTCCGACAGCTCACAGAACATATCCAGGAGTGTGGGCCCTTGTGATCTG 1656
QY 481 ValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500
Db 1657 GTGATTTGGGGAGTCTCTCAATGACTCTTCAATCTCAACCTGTGCTCGCAAGGCTTC 1716
QY 501 TyrGluGlyThrGlyArgLeuPheGluPheGluPheTyrArgLeuLeuHisAspAlaArgPro 520
Db 1717 TACAGGGCACTGGCGGCTCTTCTTGAATTCACCGCTCTCGCATGATGCGGGGCC 1776
QY 521 LysGluGlyAspAspArgProPhePheTrrPheLeuPheGluAsnValValAlaMetGlyVal 540

```

Dh 1777 AAGAGAGATGATGATGCGCCCTTCTTGTGGCTTTTGAGATGTGGTGGCCATGGCGTT 1836  
Qy 541 SerAspLysArgAspLysSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
Dh 1837 AGTGACAAAGAGGACATCTCGCATTTCTCGATCCCAACCTGTGATGATGCCAA 1896  
Qy 561 GluValSerAlaIleHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg 580  
Dh 1897 GAAGGTCTCAGCTCACACAGGCGCCGCTACTTCTGGGGGTAACTTCCGGTATGAACAGG 1956  
Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGlnHisGlyArg 600  
Dh 1957 CCGTTGGCATCCCATCTGTGAATGATAGCTGGAGCTGAGAGAGTGTCTGGAGCATGGCAGG 2016  
Qy 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
Dh 2017 ATAGCCCAAGTTCAGCAAGAGGAGACCATTAACGAGTCAAACTCCATTAAGACAGGCG 2076  
Qy 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGlu 640  
Dh 2077 AAGACACAGCATTTCTCTGTTCATGATGATAAGAAAGAGACATCTTATGTGCACCTGAA 2136  
Qy 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
Dh 2137 ATGGAAAGGATATTGTGTTCCAGTCCACTATACAGACCTTCCAACATGAGCCGCTTG 2196  
Qy 661 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
Dh 2197 GCGAGGCAAGAGACTGCTGGGCGCGGTCAATGAGAGCGTGCACAGTCAACGCTTTCGCT 2256  
Qy 681 ProLeuLysGluTyrPheAlaCysVal 689  
Dh 2257 CCGCTGAAGAGATTTTGGCTGTGTG 2283

RESULT 2  
LOCUS CS050243 3005 bp DNA linear PAT 23-MAR-2005  
DEFINITION Sequence 27 from Patent WO2005021757.  
ACCESSION CS050243  
VERSION CS050243.1 GI:61889498  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1.  
REFERENCE  
AUTHORS Tome, P. H. and van Rompaey, L.  
TITLE Polypeptides and polynucleotides for use as a medicament  
JOURNAL Patent: WO 2005021757-A 27 10-MAR-2005;  
Galapagos Genomics N.V. (BE)  
FEATURES  
source  
1.3005  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 3005  
Score: 3763.00 Matches: 689  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0

US-10-623-813-86 (1-689) x CS050243 (1-3005)

Qy 1 MetAsnAlaValGluGluLeuGlnGlyProGlyLysSerLysValGluGluAlaSer 20  
Dh 899 ATGAAATGCTGTGGAAGAAACAGAGGCGCCGGGAGTCTCAGAAAGTGTGGAGGCGCAGC 958  
Qy 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrTrpGlu 40

Dh 959 CCTCTGCTGTGACAGAGCCCACTGACCCCGCATCCCCACTGTGGCTACCAAGCCTGAG 1018  
Qy 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
Dh 1019 CCGGTGGGGTCCGATGCTCGGTGGGACAAAGATGCCAACAAAGCAGGCGATGACGAGCAGAG 1078  
Qy 61 TyrGluAspGlyValArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuAspGlyPhe 80  
Dh 1079 TAGAGAGAGCGCGGGGCTTTGGCATTTGGAGCTGTGTGGGGAACCTGCCGGGCTTC 1138  
Qy 81 SerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
Dh 1139 TCTGTGTGCGAGCGCGCATTTGTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1198  
Qy 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys 120  
Dh 1199 GGCACCCGCTGGGTCACTGTGTTCGAGACGCGCAATTCACGTGTGTGTGTGTGTGTGTGT 1258  
Qy 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140  
Dh 1259 CTGATGCTCGTAGCTCGTTTGTGCACTGCTTCACACAGGCCACGTAACAACAGCAGCCC 1318  
Qy 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
Dh 1319 ATGTACCGCAAGCCATCTACAGAGTCTGCAAGTGGCCAGACGCCGCGGGGGAAGCTG 1378  
Qy 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValAlaGlnAsn 180  
Dh 1379 TTCCCGGTGTGCACGACGACGATGAGTGAACACGCCAAGGCCGTGGAGGTGCAGAAC 1438  
Qy 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
Dh 1439 AAGCCATATGTAATGAGGCGCCCTGGGGGCTTCCACCTTCTGGCCCTTAAGGCGCTGGAG 1498  
Qy 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro 220  
Dh 1499 CCACGAGAAAGAGAGAAATCCCTACAAAGATGACACCGACATGTGGGTGGAACCT 1558  
Qy 221 GluAlaAlaIleTyrAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
Dh 1559 GAGGACGCTGCTACCGACACCTCCACCGACCCCAAAACCCCGGAAAGACAGCGGAG 1618  
Qy 241 LysProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluVal 260  
Dh 1619 AAGCCCAAGGTCAAGAGATTTATGATGAGCGCACAGAGAGCGCTGTGTACGAGGTG 1678  
Qy 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
Dh 1679 CCGCAGAAATGCGCGAATCTTGAAGACATCTGCATCTCTGTGGAGCGCTCAATGTTACC 1738  
Qy 281 LeuGlnHisProLeuPheValGlyLysCysGlnAsnCysLysAsnCysPheLeuGlu 300  
Dh 1739 CTGGAACACCCCTCTCTGTGGAGGAATGTGCCAAATGTCCAAAGAACTGCTTTCGAG 1798  
Qy 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
Dh 1799 TGTGCTACCAAGACACACAGCGGCTACAGTCTTACTGACCATCTGTCTGTGGGGC 1858  
Qy 321 ArgGluValLeuMetCysGlyLysAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
Dh 1859 CGTGAGTGTCTCATGTGCGGAAACAAACATCTGTGAGAGTGTTCGCGGAGTGTGTG 1918  
Qy 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspProTrpAsnCys 360  
Dh 1919 GACCTCTTGTGTGGGCGCGGGGCTGCCAGGACCATTAAGGAACCCCTGGAACCTGC 1978  
Qy 361 TyrMetCysGlyHisLysGlyThrTyrGlyLysLeuLeuArgArgArgLysAspTrpProSer 380  
Dh 1979 TACATGTGGGGCACAAAGGTACCTACGGGCTGCTGCGCGGAGAGAGACTGGCCTCC 2038  
Qy 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400

Db 2039 CGGCTCAGAGTCTTGGCTAATACCAACGACGAAATTTGACCTCCAAAGCTTAC 2098  
Qy 401 ProProvalProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
Db 2099 CCACCTGTCGCCAGCTGAGAGAGAGAACCCATCCGGGTGCTCTCTCTTGATGAAATC 2158  
Qy 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleIleAsp 440  
Db 2159 GCTACAGGGCTCTCTGCTGTAAGACTTGGGCAATTCAGGTGACCGGCTACATTCCTCG 2218  
Qy 441 GluValCysGluAspSerIleThrValGlyMetCValArgHisGlnGlyLysIleMetTyr 460  
Db 2219 GAGGTGTGTGAGATTCATCATCGGTGGCAAGTCCGGCACACAGGGAGATCTATGAC 2278  
Qy 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGlyTyrGlyProPheAspLeu 480  
Db 2279 GTCCGGGACGTCGCCAGCGTCAACAGAAAGCATATCCAGAGTGGGCCCATTCATCTG 2338  
Qy 481 ValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
Db 2339 GTGATTTGGGGCGACGCTCTGCATATGACCTTCATGTCACCTGCTCGCAAGGGCTC 2398  
Qy 501 TyGlnGlyTyrGlyArgLeuPheGluPheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
Db 2399 TACGAGGCACTGGCGGCTCTTTCTTGAGTTCTACCGCTCTCTCATGATCGCGGCC 2458  
Qy 521 LysGlnGlyAspAspArgProPhePheTyrLeuPheGluAsnValValAlaMetGlyVal 540  
Db 2459 AAGGAGGAGATGATGCGCCCTTCTTGCGCTCTTGAGAAATGTGTGGCCATGGGCGTT 2518  
Qy 541 SerAspLysArgAspIleSerArgPheLeuGlnSerAsnProValMetIleAspAlaLys 560  
Db 2519 AGTACCAAGAGGAGCATTCGGGATTTCTCGAGTCCAACTCTGATATATGATGCCAA 2578  
Qy 561 GluValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArg 580  
Db 2579 GAAGTGACGTGCACACAGGGCCCGCTACTTCTGGGTAACCTTCCGGTATGAAACGG 2638  
Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGlnGlyCysValLeuGluHisGlyArg 600  
Db 2639 CCGTTGGCATCCACTGTAATGATAGCTGAGCTGAGAGTGTCTGAGCATGCGCAGG 2698  
Qy 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
Db 2699 ATAGCAAGTTCAGCAAGAGTGAGACCACTACTACAGGCTCAACTCCATTAAGCAGGAC 2758  
Qy 621 LysAspGlnHisPheProValPheMetAsnGlnLysGluAspIleLeuTyrCysThrGlu 640  
Db 2759 AAAGACCAACATTTTCTGTCTTCAATGATGAAAGAGGACATCTTATGGTGCACCTGA 2818  
Qy 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
Db 2819 ATGGAAGAGGATTTGGTTCCCACTCCACTATCTGACGCTCCCAACATGAGCGGCTTG 2878  
Qy 661 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
Db 2879 GCGAGGCGAGACGCTGCGCGGCTCATGAGCGGTCCAGATCATCCGCCACTTTCCT 2938  
Qy 681 ProLeuLysGluTyrPheAlaCysVal 689  
Db 2939 CCGCTGAAGGAGTATTTGGCGTGTG 2965

RESULT 3  
LOCUS AF067972 3005 bp mRNA linear PRI 12-FEB-2001  
DEFINITION Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,  
complete cds.  
ACCESSION AF067972.2 GI:12746531  
VERSION AF067972.2  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
FEATURES  
source  
1. .3005  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2p23; between D2S171 and D2S174; 48-50 cM"  
1. .3005  
/gene="DNMT3A"  
230. .2968  
/gene="DNMT3A"  
/function="de novo DNA cytosine methyltransferase"  
/product="DNA cytosine methyltransferase 3 alpha"  
/protein\_id="AAD3084.2"  
/db\_xref="GI:12746532"  
/translation="MPAPSSGPGDPTSSAAAREEDRKDGEBOEPGRKEORPSTT  
ARKVGRGRKRPVYSSGPTPPDPATVTSKPSMAOOSGASSELLPNDLEKRSPOE  
EGSPAGQKGAPEBGBABETLPEARAVENGCTTKEGRGAPEAGKEKENTIS  
MKWESGRLRGIGWESSLROBPMPPLTVOADPYITSRKDEWLARKRENAEK  
KVIAGMAVENOGPGESQVEBASPPAVOQPDPAATPEPAAGSDAGDNATK  
AGDEPEYEDGREGFISGLVWKLKRGSMGRIIVSMWGRSALAGTVMVFGD  
KFSVYEXKIMPLSPCSAFAPQATYNKQMPYRAIYVLOVASRAKLFPVCHDS  
SDPAKAVYONKMIEMALGROPBGKGLPEPEENPKVAYTDMVPEAAAYAP  
PPEAKPRKSTAEKPKVETIIDERTRLYEVROKRNIEDICISGSLNTLEHL  
FVGMCQNCNCELECAVQYDDGQYOSCTICCGREVLKCGNNCCRCFCVBCVDL  
VCPGAQAPAKIEDPWNCMYGHKGTGLLRREBPRLQMFANNHDOERDPKVP  
PVPAEKRPKRLVSLFPGIATGLVLDLGIQVDRYIASEVCEBSITVGMRHQKIM  
YVGDVRSYOKHIOEMGPPDLVIGGSPCNLSIVNPAKGLYEGLRLFEEFYLLHD  
ARKREGDREFPPLFENVVAMGVSDKRDISRLPESNPVMTDAKEVSAHRAARYWGL  
PGNMRPLASTVNDLEIQLCEHGRILAKSKVRIITTRRSIKQKQHPVPFNWKE  
DILMTEMERVFQPPVHYTVDSNMSRLARQLLRGSVSPVIRHLPALKEFYACV"

ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 3005  
Score: 3763.00 Matches: 689  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 5 Gaps: 0

US-10-623-813-86 (1-689) x AF067972 (1-3005)

Qy 1 MetAsnAlaValGlnGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAlaSer 20  
Db 899 ATGAATGCTGTGGAAGAAACAGAGGGCCCGGGGAGTCTCAGAAAGTGGAGAGCCAGC 958  
Qy 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaTatTrtThProGlu 40  
Db 959 CTTCTCTGTGCGAGCGCCCACTGACCCCGCATCCCCCACTGTGCTACCAACCCCTGAG 1018

QY 41 ProValGlySerAspAlaGlyAspLeuAspAlaThrValAlaGlyAspAspGluProGlu 60  
 DB 1019 CCGGTGGGGTCCGATGCTGGGAGACAGAAATGCACCAAGACGGCATGACAGCCAGAG 1078  
 QY 61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPhe 80  
 DB 1079 TAGAGAGACGGCGGGCTTTGGCATTGGGGAGCTGGGTGGGGAACATGCGGGGCTTC 1138  
 QY 81 SerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
 DB 1139 TCTGTGGGCGCAGGCCCATTTGTCTTGTGGATGACGGGCCGGAACCGAGCGCTGAA 1198  
 QY 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys 120  
 DB 1199 GGCAACCGCTGGGATGATGGGTTCGAGACGGCAAAATTCACATGGTGTGTGTTGAGAG 1258  
 QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGluPro 140  
 DB 1259 CTGATGCCGTGAGCTGTTTTCAGTGGCGTCCACAGCGCCACGTACAAACAGCACCCC 1318  
 QY 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 DB 1319 ATGTACCGCAAGCCATCTACAGGTCTCTGAGGTGGCCAGACCGCGGGGAAACCTG 1378  
 QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValAsn 180  
 DB 1379 TTCGGGTGTGCACGACAGCGATGAGGTGACCTGCCAAGCCGTGGAGGTGCAGAAC 1438  
 QY 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
 DB 1439 AACCCATGATTGAATGGCCCTGGGGGGCTTCAGGCTTCGCGCCCTAAGGGGCTGGAG 1498  
 QY 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro 220  
 DB 1499 CCACCGAAGAAAGAAAGAAATCCCTCAAGAGTGTACACGCAATGTGGTGAACCT 1558  
 QY 221 GluAlaAlaIleAlaTyrAlaProProProProAlaLysLysProAlaGlyLysSerThrAlaGlu 240  
 DB 1559 GAGGACAGCTGCTACGACCACTCCACAGCCAAAGAGCCCGGAAAGACACGCCGAG 1618  
 QY 241 LysProLysValLysGluIleIleAspGluArgThrArgLysLeuValTyrGluVal 260  
 DB 1619 AACCCCAAGTCAAGAGGATTATTGATGAGGCCACAGAGCGGGCTGGTACAGAGTG 1678  
 QY 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 DB 1679 CGGCAGAAATGCCGGAACATTGAGGACATCTGCATCTCTGTGGGACCTCAAGTACC 1738  
 QY 281 LeuGluHisProLeuPheValGlyLysMetCysGlnAsnLysCysLysAsnLysPheLeuGlu 300  
 DB 1739 CTGGAAACCCCTCTTCTGTGGAGATGTGCCAAATGCAAGACTGCTTTCTGGAG 1798  
 QY 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
 DB 1799 TGTGGCAGACAGTACGACGAGAGCGGTACAGTCCACTGACATCTGTGTGGGGG 1858  
 QY 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
 DB 1859 CGTGAGGTGCTCATGTGCGGAAACAAACATGCTGCGAGTGCTTTCTGTGAGTGTG 1918  
 QY 341 AspLeuLeuValGlyProGlyValAlaGluAlaIleLysGluAspProTyrAsnLys 360  
 DB 1919 GACCTCTGGTGGGGCGGGGGGCTGCCAGGCAACCAATTAAAGAAAGCCCTGGAACTGC 1978  
 QY 361 TyrMetCysGlyLysLysGlyThrTyrGlyLeuLeuArgArgArgGluAspTrpProSer 380  
 DB 1979 TACATGTGGCGGACAGAGGTAACCTACGGGTGTGGGGGGCAGAGGACTGCCCTCC 2038  
 QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400  
 DB 2039 CGGCTCCAGATGTTCTTCGCTAAATTAACACAGCACAGAAATTTGACCTCCAAAGTTTAC 2098  
 QY 401 ProProValProAlaGlyLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420

DB 2099 CCACTGTCCCACTAGAAAGAGAGGCCATCCGGGTGCTGTCTCTTGTGATGAAATC 2158  
 QY 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAlaAspArgTyrIleAlaSer 440  
 DB 2159 GCTAACAGGGCTCTCTGTGTGGAAGGACTTGGGCAATTCAGGTGACCGCTACATTTGCC 2218  
 QY 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
 DB 2219 GAGGTGTGAGGACTTCATACCGTGGCATGTGGTGGGACCAAGGAAAGATCATTTAC 2278  
 QY 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnLysTrpGlyProPheAspLeu 480  
 DB 2279 GTCGGGGAGCTCCGACGCTCACAGAGCATATCAGAGAGTGGGCCCATTCGATCTG 2338  
 QY 481 ValIleGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 DB 2339 GTGATTTGGGGCAGTCCCTGCATATGACCTTCATGTCAACCTGCTCCAAAGGGCTC 2398  
 QY 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuHisAspAlaArgPro 520  
 DB 2399 TACGAGGACCTGGCCGGCTCTTTCAGTTTACCGCTCTTCGATGATGGCGGCCCC 2458  
 QY 521 LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValAlaAlaMetGlyVal 540  
 DB 2459 AAGAGGGAGATGATGCGCCCTTCTTGCTCTTGAGAAATGTGGTGGCCATGGCGCTT 2518  
 QY 541 SerAspLysArgAspIleSerArgPheLeuGlnSerAsnProValMetIleAspAlaLys 560  
 DB 2519 AGTGACAGAGGAGACATCTCGCATATTCGAGTCCAAACCTGTGATGATTAAGCCAAA 2578  
 QY 561 GluValSerAlaIleHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg 580  
 DB 2579 GAAGGTCACTGCACACAGAGCCCGCTACTTCTGGGTAACTTCCGGTATGAACAGG 2638  
 QY 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnLysCysLeuGluHisGlyArg 600  
 DB 2639 CGTTGGCATCCACTGGAATGATTAAGCTGAGCTGACGAGAGTGTGGAGCATGGCAGG 2698  
 QY 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
 DB 2699 ATAGCCAAAGTTCCAGAAAGTGAAGACCATTAATCAAGAGTCCCAATTAAGCAGGGC 2758  
 QY 621 LysAspGluHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGlu 640  
 DB 2759 AAAGACAGCAATTTCTGTCTTCAATGATGAGAAAGAGACATCTTATGTGCATCGAA 2818  
 QY 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
 DB 2819 ATGGAAGGGTATTGTTGCCAGTCCACTATATGACGTCTCCAAATGAGCGGCTTG 2878  
 QY 661 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
 DB 2879 GCGAGGACAGAGCATGTGGCCCGTGCATGAGCGTGCACATTCGCGCACCTTCGCT 2938  
 QY 681 ProLeuLysGluTyrPheAlaCysVal 689  
 DB 2939 CCGCTGAAGGATTTTGTGCTGTG 2965

RESULT 4  
 CS050242 3604 bp DNA linear PAT 23-MAR-2005  
 LOCUS  
 DEFINITION Sequence 26 from Patent WO2005021757.  
 ACCESSION CS050242  
 VERSION CS050242.1 GI:61889497  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 REFERENCE 1  
 AUTHORS Tomme, P. H. and van Rompaey, L.

TITLE polypeptides and polynucleotides for use as a medicament  
JOURNAL Patent: WO 2005021757-A 26-10-MAR-2005;  
Galapagos Genomics N.V. (BE)  
FEATURES  
source Location/Qualifiers  
1. .3604  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 3604  
Score: 3763.00 Matches: 689  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0  
US-10-623-813-86 (1-689) X CS050242 (1-3604)  
QY 1 MetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAlaSer 20  
DB 217 ATGAATGCTGTGGAGAAACCAAGGGCCCGGGAGTCTCAGAAAGTGGAGAGCCAGC 276  
QY 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40  
DB 277 CCTCTGCTGGAGAGAGCCAGCTGACCCGATCCCTGCTGGCTTACCAAGCTTGGAG 336  
QY 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
DB 337 CCGGTGGGTCTCGATGCTGGGGACAAAGATGCAACCAAGCAGCGATGACGAGCCAGAG 396  
QY 61 TyrGluAspGlyValGlyGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgLysPhe 80  
DB 397 TACGAGAGCGCCCGGGCTTTGGCATTTGGGAGCTGTGTGGGAAACTGCGGGGCTTC 456  
QY 81 SerTrpTrpProGlyValGlyLeuAlaSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
DB 457 TCTGTGTGGCCAGGCGCGCATTTGTCTTGTGGATGAGAGGCGCGGACCGAGAGCTGAA 516  
QY 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValLysValGluLys 120  
DB 517 GGCAACCGGTGGGTCAATGTGGTTCGAGACCGCAAAATTTCTCAGTGTGTGTGAGAGAG 576  
QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyraAsnLysGluPro 140  
DB 577 CTGATGGCCGTGACCTCGTTTGGAGTGCCTTCCACAGCGCCAGCTACCAAGCAGAGCC 636  
QY 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
DB 637 ATGACCGCAAGGCATCTTACGAGGTCTCTGACGTGGCCAGCAGCCCGCGGGAACTG 696  
QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsn 180  
DB 697 TTCCCGGTGTGCCAGCAGAGATGAGATGACACTGTCCAAGCGCTGGAGGTGCAGAAC 756  
QY 181 LysPrometIleGluTyrPalaLeuGlyGlyPheGlnProSerArgLysProLysGlyLeuGlu 200  
DB 757 AAGGCCATGATGAATGGCCCTGTGGGGCTTCCAGGCTTCTGCGCTTAAGGGCTGGAG 816  
QY 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro 220  
DB 817 CCACAGAAAGAAAGAAAGATCTCTTACAAAGATGTAACGGAACATGTGGTGAACCT 876  
QY 221 GluAlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
DB 877 GAGGCACTGCTGACGACCACTCCACACAGCAAAAGCCCGGAGAGAGACAGCCGAG 936  
QY 241 LysProLysValLysGlyIleIleAspGluArgThrArgLysGluLeuValTyrGluVal 260  
DB 937 AAGGCCAAGGTCAAGGAGATTATGATGAGCGCAACAAGAGCGGCTGGTGTACGAGGTG 996  
QY 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280

DB 997 CGGCAGAGAGCGCGGAACATTTAGGACATCTGCATCTCTGTGGAGCCCAATGTTACC 1056  
QY 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
DB 1057 CTGGAACACCCCTCTTCTGTTGAGGAAATGTGCCAAACGTGCAAGAACTCTTCTTGAG 1116  
QY 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
DB 1117 TGTGCTTACAGTACGACGACGAGCGGCTTACAGTCTTACCTGACCATCTGCTGTGGGGC 1176  
QY 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
DB 1177 GGTGAGGTGCTCATGTGGCGGAAACAACTGCTGAGGTGCTTTTGGTGGAGTGTG 1236  
QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyrPheCys 360  
DB 1237 GACCTCTTGTGGGGCGGGGGCTGGCCAGGAGCCATTAAAGAAAGCCCTGGAACTGC 1296  
QY 361 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTrpProSer 380  
DB 1297 TACATGTGCGGACAAAGGTACTTACGGGCTGCTGGCGGCGAGAGACTGGCCCTCC 1356  
QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400  
DB 1357 CGGCTCAGATGTTCTTCGCTAATTAACAGACAGCAAGAAATTTGACCTCCAAAGTTTAC 1416  
QY 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
DB 1417 CCACTGTCTCCACCTGAGAGAGAGAGCCCATCGGGTGTGTCCTCTTGTGATGAAATC 1476  
QY 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
DB 1477 GCTACAGAGGCTCTGTGTGCTGAAGAGACTTGGGCATTCAGGTGACCGCTCATTTGCTCG 1536  
QY 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
DB 1537 GAGGTGTGAGAGATCCATCACGATGGCATGTGTCGAGCACAGGGAGAAAGATCATGTAC 1596  
QY 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeu 480  
DB 1597 GTGCGGAGCTCGCAGCGTCACAGAGACATATCCAGGAGTGGGGCCCATTCGATCTG 1656  
QY 481 ValIleGlyLysProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
DB 1657 GTGATTTGGGGGACGTCCCTGCATTAAGCTCTCATGCTCAACCTGTCGAAGGGCTTC 1716  
QY 501 TyrGluGlyThrGlyValArgLeuPheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
DB 1717 TACGAGGGCACTGGCGGCTCTTCTTGAATTTACCGCTCTGCAATGATGGCGGCCCC 1776  
QY 521 LysGluGlyAspAspArgProPhePheTrpLeuPheGluLysAsnValAlaIleMetGlyVal 540  
DB 1777 AAGAGAGGAATGATGCCCCCTTCTTGTGCTTGTGAATGATGGTGGCCATAGGGGCTT 1836  
QY 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
DB 1837 AGTGACAAAGGAGACATCTGCGATTTCTCGAGTCCAACTCTGATGATGATGCCAAA 1896  
QY 561 GluValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArg 580  
DB 1897 GAAGTGCAGCTCACACAGGGCCCGCTACTTCTGGGGTAACTTCCCGGTATGAACAG 1956  
QY 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArg 600  
DB 1957 CCGTGGCATCTCCATGTAATGATTAAGCTGAGCTGAGAGAGTGTGGAGCATGGCAGG 2016  
QY 601 IleAlaLysPheSerLysValArgTrpIleThrThrArgSerAsnSerIleLysGlnGly 620  
DB 2017 ATAGCCAAAGTTACGAAAGTGAAGACATTAACGAGGTCAAACTCATTAAGCAGAGGC 2076  
QY 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGlu 640

Db	2077	AAAGACGACATTTCTGCTTCATGATGAGAAAGAGCATCTTATGTCACGTAA	2136
Qy	641	MetGluArgValPheGlyPheProValHisTyrThrSerValSerAsnMetSerArgLeu	660
Db	2137	ATGGAAGGGTATTTGGTTTCCAGTCCACTATACACGCTTCCACATGAGCGGCTTG	2196
Qy	661	AlaArgGlnArgLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAla	680
Db	2197	GCGAGGACAGACTGCTGGGCGGTCATGAGCGTGCACATCCACCTCTTGCT	2256
Qy	681	ProLeuArgGlyTyrPheAlaCysVal	689
Db	2257	CCGCTGAAGAGTATTTGGCTGTGTG	2283
RESULT 5			
AF331856		4258 bp	linear
LOCUS			
DEFINITION	Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,		
ACCESSION	AF331856		
VERSION	AF331856.1	GI:18033252	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 4258)		
AUTHORS	Ni,J., Pradhan,S. and Roberts,R.J.		
JOURNAL	Cloning, expression and characterization of human DNMT3 genes		
REFERENCE	2 (bases 1 to 4258)		
AUTHORS	Ni,J., Pradhan,S. and Roberts,R.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,		
REFERENCE	Beverly, MA 01915, USA		
FEATURES			
source	1..4258	Location/Qualifiers	
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="2"		
	1..4258		
gene	/gene="DNMT3A"		
CDS	227..2956		
	/codon_start=1		
	/product="DNA cytosine methyltransferase 3 alpha"		
	/protein_id="AA57039.1"		
	/translation="MSSSGRGDTSSSAAREEDRKQGEBOEPGRKEERQEPSTARK		
	VGPFGKRRKHPVPSGDTKPDPAVISKSPWADSGASSELPLPGDLEKSEPEPES		
	PAGGQGGAPAEAGEGAETLPEASRAVENCCPTKGRGAPAGKEQKTNIESKM		
	EGSGRGLRGGLGWESSLRQPMPLTFOADPYIISKRKDEWLAKREAEKAVYI		
	AGNAAVEENQGESQKEASPPAOOPDPAFTVAATTPBEVGSADGKNTKAGD		
	DEPEYDNGGFGTGLVWGKLRGSPWMPGIVSMATTPGRABRGIRWMMFGDGKS		
	VVCVEKLMLPSFCSAFHQITYKKQPMRYALIEVIOVASRAGKLPRVCHDSBDT		
	AKAVEVONKPEIEMALJGFQPSGPGKGLPEPEEKPKVEYTDWMPEEAATAPPP		
	AKPRSTAEKPKVKEIIDERTRERLVYERQKRNIEDICISGSLNVLLEHPLVG		
	GMAQONKCNFLECAVYDDGYQSYCTICCGREVLVLMGNNCCRCFCECVLLVGP		
	GAQAALIKEDPMNCGHKGKTYGLRLRRRDMPSRLQMFPAHHNDOEPKYPYVP		
	AEKKRIYLSLFDGIATGLILVLDGIQDVRVIASEVGDSTIVMVRHQGLMVG		
	DVRSVTOKHIOEMGFDPDLVIGSPCNDLSVNPARKGLVGTGRLLPFERILLHDAP		
	KGGDDPFPFMLENVAVMGVSDKDISRFLESNPVWIDAEVSAARHARTFNGNLGM		
	NRPALSTVNDKLEQCLEHGRILAKFSKVTITTRSNSIKQGDQHFPFVFNKEDIL		
	WCTEMERVFGFPVHYTVDVSNMRLARQLLGRSWSVPVIRHLPAPLKEFYACV"		
ORIGIN			
Alignment Scores:			
Score:	0	Length:	4258
Percent Similarity:	3763.00	Matches:	689
Best Local Similarity:	100.0%	Conservative:	0
	100.0%	Mismatches:	0

Query Match:	100.0%	Indels:	0
DB:	5	Gaps:	0
US-10-623-813-86 (1-689) x AF331856 (1-4258)			
Qy	1	MetAsnAlaValGluGluGlnGlnGlyProGlyGluSerGlnLysValGluGluAlaSer	20
Db	887	ATGAATGCTGTGAAGAAAACAGAGGGCCCGGGAGCTCAGAAAGGTGAGAGGCCAGC	946
Qy	21	ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrProGlu	40
Db	947	CCCTCGCTGTGCAGAGCCGACCTGACCCCGCATCCCACTGTGTACAGCCCTGAG	1006
Qy	41	ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu	60
Db	1007	CCCGTGGGTCCATGCTGGGACAAAGATGCCACCAAGCAGGCTGACGAGCCAGAG	1066
Qy	61	TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPhe	80
Db	1067	TACGAGAGCGCGGGGCTTTGGCATTTGGGAGCTGTGTGGGGAAACTGCGGGGCTTC	1126
Qy	81	SerTyrTyrProGlyArgGlyLeuValSerTyrTyrMetThrGlyArgSerArgAlaAlaGlu	100
Db	1127	TCTGTGGCCAGCGCCGCAATGTGTCTGTGTGATGACGGCCGAGCCGAGAGCTGAA	1186
Qy	101	GlyThrArgTyrValMetTyrPheGlyAspGlyLysPheSerValValCysValGluLys	120
Db	1187	GACACCCGCTGGGTCTATGTGTGTCGAGACGCGCAATTCCTCATGTGTGTGTGAGAG	1246
Qy	121	LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro	140
Db	1247	CTGATCCGCTGAGCTGCTTTTTCAGTGGCTTCCACAGGCCAGTACCAAGCAGACCCC	1306
Qy	141	MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu	160
Db	1307	ATGTACCGCAAAACCACTTACGAGGTCTGCAAGTGGCCGAGCCGCGGGGAAACTG	1366
Qy	161	PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluAlaAsn	180
Db	1367	TTCCCGGTGTGCACACACAGCATGAGTGAAGTGAACCTGCCAAGCCGTGAGGTGCAAG	1426
Qy	181	LysProMetIleGluTyrAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu	200
Db	1427	AAAGCCATGATTGAATGAGGCGCTGGGGGGGCTTCACGCTTCTGAGGCTTGAG	1486
Qy	201	ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTyrValGluPro	220
Db	1487	CCACCGAAGAAAGAAAGATCCCTTCAAGAAAGTGTACCGGACATGTGGGTGAACTT	1546
Qy	221	GluAlaAlaAlaTyrAlaProProProAlaLysLysProArgLysSerThrAlaGlu	240
Db	1547	GAGCGAGTGTCTTACCAACACCTTCCACAGCCCAAAAGCCCGGAAGACACAGCGAG	1606
Qy	241	LysProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluVal	260
Db	1607	AAAGCCAGGTAAAGAGATTATTTATGAGCGGCAAGAAGAGCGGCTGTGTACAGGTG	1666
Qy	261	ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr	280
Db	1667	CGGCAAAAGTGCAGGAACATTGAGCAATCTGATCTCTGTGGAGGCTCAATGTTAAC	1726
Qy	281	LeuGlnHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnLysPheLeuGlu	300
Db	1727	CTGGAACACCCCTCTTCTGTGAGAGAAATGTGCCAAAATCGCAAGAACTCTTCTGAG	1786
Qy	301	CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly	320
Db	1787	TGTGCGTACCAAGTACGACGACGAGCGGCTACAGTCTTACGTCACCATCTGCTGTGGGGG	1846
Qy	321	ArgGluValLeuMetCysGlyLysAsnAsnLysCysArgCysPheCysValGluCysVal	340
Db	1847	CGTGAAGTCTCATGTGGCAAAACAACACTGTGACAGGTCTTTTGGCGTGTGTGTG	1906

QY	341	AspLeuLeuValAlGlyProGlyValAlaIaGlnAlaAlaIleLysGluAspProCTTAsCys	360
Db	1907	GACCTCTTGGTGGGGCCGGGGGGCTCCAGAGGCACTTAAGAAAGACCTCGAACTGC	1966
QY	361	TyrMeCysGlyHisLysGlyThrTyrgLysLeuAArgArgGlyAspTrpProSer	380
Db	1967	TACATGTCCGGGCACAAAGGATCTTACGGGCTGCTCGCGCGGAGAGACTGGCCCTCC	2026
QY	381	ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnLupheAspProProlYsValTyr	400
Db	2027	CCGGCTCCAGATGTTCTTGGCTATAACACAGCACAGAAATTGAACCTCCAAAGGTTAC	2086
QY	401	ProProlValProAlaGlnLysArgLysProlIleArgValLleuSerLupheAspGlyLe	420
Db	2087	CCACCTGTCCAGCTCGAAGAGGAAAGGCCATCCGGGTGCTGTCTCTTTGAAGGATC	2146
QY	421	AlaThrGlyLeuLeuValLleuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer	440
Db	2147	GCTACAGGGCTTCCTGTGCTCTGAAGACTTGGGCACTTCAAGGTGAGACCTTACATTGCC	2206
QY	441	GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr	460
Db	2207	GAGGTGTGTAGGAACTCATCATCGGTGGCATGGTGGCGGCACAGGGAGAAATCATGTAC	2266
QY	461	ValGlyAspValAArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeu	480
Db	2267	GTCGGGGAGCTCCCGACGCTCACACAGAAAGATATCCAGAGTGGGGCCCATTCGATCTG	2326
QY	481	ValIleGlyLysSerProCysAsnAspLeuSerIleValIaAsnProAlaArgLysGlyLeu	500
Db	2327	GTGATTTGGGGGAGTCCCTCGCAATACCTCTCATCGTCAACCTCTGCTCGAAGGGCTC	2386
QY	501	TyrgLuglyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro	520
Db	2387	TACAGAGGCACCTGGCGGCTCTCTTTAGTTTACCGGCTCTGTCAGATAGCGGGGCC	2446
QY	521	LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyVal	540
Db	2447	AAGGGAGAGATGATCGCCCTCTCTTCTGGGCTTTGGAATGTGGTGGCATGGGCTT	2506
QY	541	SerAspLysAspAspIleSerArgPheLeuGlnLysAsnProValMetIleAspAlaLys	560
Db	2507	AGTGCAAAGAGGACATCGGGAATTTCTGAGTCCAACTCTGTATGATGATGCCAA	2566
QY	561	GluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg	580
Db	2567	GAAGTGTACGTGCACACAGAGGCCCGGCTACTTCTGGGTAACTCTCCGGTATGAACAG	2626
QY	581	ProLeuAlaSerThrValAsnAspLysLeuGlnLysCysLeuGlnHisGlyArg	600
Db	2627	CCGTTGGCATCCACTGTAATGATGAAGTGGAGGCTCAGAGATGCTTGGACATGGCAGG	2686
QY	601	IleAlaLysPheSerLysValArgThrIleThrTrhArgSerAsnSerIleLysGlnGly	620
Db	2687	ATAGCAAACTTCAGCAAAAGTAGAGGCCATTACTACAGAGTAAACTCCATTAACAGAGGC	2746
QY	621	LysAspGlnHisPheProValPheMetAsnGlnLysGluAspIleLeuTrpCysThrGln	640
Db	2747	AAAGACCAACATTTCTCTGTCTCTCATGAATAGAAAGAGACATTTATGTTGCACTGAA	2806
QY	641	MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu	660
Db	2807	ATGGAAGAGGATATTGTTTCCCACTCACAATCTGACGCTCCCAACATAGCCGGCTTG	2866
QY	661	AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla	680
Db	2867	GCGAGGCAAGAACTGCTGGGCCGGGTCATGAGAGCGTGCATATCCGCACTTTCGCT	2926
QY	681	ProLeuLysGluTyrPheAlaCysVal	689
Db	2927	CCGCTGAAGAGATATTTTGGCTGTCTG 2953	

[illegible]

QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaValAlaGluValAsn 180  
 Db 1387 TTCGGGGTCCACGACGAGTATGAGTACACTGCGCAAGCCGTGAGGTGAGAAC 1446  
 QY 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProGlyGlyLeuGlu 200  
 Db 1447 AAGCCCATGATTGAATGGGCCCTGGGGGGCTTCCAGCCTTCTGGGCCCTTAAGGCCCTGAG 1506  
 QY 201 ProProGluGluGluValAsnProTyrLysGluValTyrThrAspMetTyrAlaGluPro 220  
 Db 1507 CCACGACAGAGAGAGAAATCCCTACAAAGAGTACACGACATGTGGTGGAACT 1566  
 QY 221 GluAlaAlaAlaTyrAlaProProProProAlaLysLysProAlaGlySerThrAlaGlu 240  
 Db 1567 GAGGACAGCTGCTTACGACACCTCCACGACCAAAAAGCCCGGAAAGACACAGCGAG 1626  
 QY 241 LysProLysValLysGluIleIleAspGluArgThrArgLysValTyrGluVal 260  
 Db 1627 AAGCCCAAGGTCAAGAGATTTATGATGAGCGCACAAAGAGCCGCTGGTATGAGAGTG 1686  
 QY 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 Db 1687 CGGCGAAGTCCCGGAACATTGAGGACATCTGCACTCTCTGGGAGCGCTCAATGTATACC 1746  
 QY 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
 Db 1747 CTGGAAACACCCCTCTTCTGTGGAGGAATGTGCCAAAATGCAAGAACTGCTTCTTGAG 1806  
 QY 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
 Db 1807 TGTGCGTACCAATGACAGAGAGAGCGCTACCACTCTCACTCACTCTGCTGGGGGC 1866  
 QY 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
 Db 1867 CGTGAGGTGCTCAATGTGCGGAAACAACTGCTGAGGTCTTTGCTGAGTGTGTG 1926  
 QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyrAsnCys 360  
 Db 1927 GACCTCTTGGGGGGCCGGGGGGCTGCCAGGACCCATTAAAGAAAGCCCTGGAACCTG 1986  
 QY 361 TyrMetCysGlyHisLysGlyTyrTyrGlyLeuLeuArgArgGlyGluAspTyrProSer 380  
 Db 1987 TACATGTGCGGCGCAACAGGTAACCTTACGGGCTGCTGGCGCGCAGAGAGACTGGCCCTCC 2046  
 QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400  
 Db 2047 CGGCTCCAGATGTTCTTCGTAATACACAGACAGAAATTTGACCTCCCAAGGTTTAC 2106  
 QY 401 ProProValProAlaGluValArgLysProIleArgValLeuSerLeuPheAspGlyTyr 420  
 Db 2107 CCACCTGTCCAGCTGAGAGAGAGAGAACCCATCGGGGTGCTCTCTTGTGATGAGATC 2166  
 QY 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
 Db 2167 GCTACAGGGCTCTCTGGTCTGTAAGAGACTTGGGCATTCAGGTGACCCCTCATTTGCTCG 2226  
 QY 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
 Db 2227 GAGGTGTGTAGAGCATCATCACTGAGGAGTGTGTGGCGCACAGGGAAGATCATGTAC 2286  
 QY 461 ValGluValAspValArgSerValThrGlnLysHisIleGlnGluTyrProPheAspLeu 480  
 Db 2287 GTGGGGGACCTCCGACGCTCACCAAGAGATVTCAGAGAGTGGGGCCCATTCGATCTG 2346  
 QY 481 ValIleGlyLysProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db 2347 GTGATTTGGGGGCACTCCCTGCAAGACTCTCCATCTGCAACCTGCTCCGAGGGCTC 2406  
 QY 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
 Db 2407 TACGAGGAGCACTGGCGGCTCTTCTTGATGATTACGCGCTCCGATGATGCGGCGCC 2466

QY 521 LysGluGluValAspAspArgProPhePheTrpLeuPheGluAsnValAlaIleMetGlyVal 540  
 Db 2467 AAGGAGGAGATGATCCGCCCTTCTTGTGCTTTTATGAATGTGGTGGCCATGAGCCTT 2526  
 QY 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
 Db 2527 AGTGACAAAGAGGAGCATCTGCGCATTTCTCGAGTCCAAACCTGTGATGATGATGCCAAA 2586  
 QY 561 GluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg 580  
 Db 2587 GAACTGCACCTGCACACAGGCGCCGCTACTTCTGGGGTAACTTCCCGGTATGAACAGG 2646  
 QY 581 ProLeuAlaSerThrValAsnAspLysLeuGluGlnGluCysLeuGluHisGlyArg 600  
 Db 2647 CCGTTGGCATTCACCTGTGAATGATTAAGCTGAGCTGCAGAGAGTGTCTGAGACATGGCAG 2706  
 QY 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
 Db 2707 ATAGCCAAAGTTCAAGCAAGTGAAGACCATTAACAGAGTCAAACTCCATTAAGCAGGGC 2766  
 QY 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGlu 640  
 Db 2767 AAAGACCAAGATTTTCTGCTTCATGATGATGAGAAAGAGACATCTTATGTGCACTGAA 2826  
 QY 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
 Db 2827 ATGGAAAGGGATTTGTTGTTCCAGTCCACTTAATCTGATCTTCAACATGAGCCGCTTG 2886  
 QY 661 AlaArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAla 680  
 Db 2887 GCGAGGCAAGACATGCTGGGCGCGGTCAATGAGCGTGCACATCCGCCACCTTCTGCT 2946  
 QY 681 ProLeuLysGluTyrPheAlaCysVal 689  
 Db 2947 CCGGTGAAGAGATTTTGGGTGTGTG 2973  
 RESULT 7  
 LOCUS BC043617  
 DEFINITION Homo sapiens DNA (cytosine-5'-methyltransferase 3 alpha, transcript variant 3, mRNA (cdna clone MGC:50948 IMAGE:6150112), complete cds.  
 ACCESSION BC043617  
 VERSION BC043617.1 GI:27694443  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 4294)  
 Strausberg, R.L., Feilgold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altshul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.R., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Schevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalske, U., Smalins, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932

REFERENCE 2 (bases 1 to 4294)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (09-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: "http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DC/DT/PTP  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada  
info@cgsc.bc.ca  
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Maason, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lisa Prambu, Parvaneh Saeedi, Jr Santos, Angelique Scherch, Ursula Skaleka, Duane Smalhus, Jeff Stoct, Miranda Teai, George Yang, Jacque Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov  
Series: IRAC Plate: 89 Row: 0 Column: 19  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES  
SOURCE Location/Qualifiers

1..4294  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:50948 IMAGE:6150112"  
/cissue\_type="Skin, melanotic melanoma."  
/clone\_lib="NIH MGC\_72"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..4294  
/gene="DNMT3A"  
/note="synonyms: DNMT3A2, M.Hs111A"  
/db\_xref="GeneID:1788"  
/db\_xref="MIM:602769"  
238..2976  
/gene="DNMT3A"  
/codon\_start=1  
/product="DNA cytosine methyltransferase 3 alpha, isoform a"

/protein\_id="AAH43617.1"  
/db\_xref="GI:2769444"  
/db\_xref="GeneID:1788"  
/db\_xref="MIM:602769"  
/translation="MPMPSSSGPGDTSAAERREDRKDGEOEPREKREORBST  
ARKYGRPKRKHPVESGDTPKDPAVLSKSPMAODGASBLPNEGLEKRESEQPE  
EGSPAGQGGAPAEAGAEETLPEASAVENGCCTPREGGAPEAGKEKEKENTES  
MKMGSRRLRGGLGWESSLRPWPRLTFOAGDPVYLSKRDDEMLRWREAEKKA  
KVIAGMAVENOGFESQKVEASPVQOQTPDPAFPVATTEPVSADGDKNATK  
AGDPEYEDRGFGIGELVWGKLRGFMWGRIVSMWTRSRRAEGTRVWMEGDS  
KEFVYCEKMLPSFCAFHQATYNNQMPYRKALIEVLQVASSRAGLTPVCHDSDE  
SDTKAVEONKPMTEALGFGFSGPGLPEPEEKPYKEVTTDMNVPEEAAAYAP  
PPPAKPKRSTAEKPKVKEIITDERRELVIEVRKCKNIDICTSCSLVNTLEHLP  
FVGQCONCKNCFLECAQYDDGYOSCTICCGREVLWGMNNCCGCFCEVDL  
VGPAAQAAIKEDPMNCCMGHKGTYLKRREDDPMSRLQMFANNHDOEDPQKVP  
VPEPEKKRPILVLSFDGIATGLVLKGLQVDRYIASEVCEDSITGVAMHOGKIN  
YVGQVSTQKIOEWGEPDLVIGSSPNDISVNPARKGLYEGTGRLFEPFVLLHD  
ARPEGDRPFPMFLPENVVAMGVSDPKRISFLSNPMDIAKESAHARRRYWGNL  
PGNMRPLASTVNDKLEIQECLEHGRIAKFSKVRITTTTSNSIKGQKDHFEVMEKE  
DILWCTEMERVFEPVHYTVDVSNMRLRRLRLGRSMSPVIRHLFALKEYFACV"

ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 4294  
Score: 3763.00 Matches: 689  
Percent Similarity: 100.0% Conserves: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
Gaps: 0

US-10-623-813-86 (1-689) x BC043617 (1-4294)

QY 1 MetasnlalavalgiugluasnginglyProclYgluseglnyvalgluglu1a1ser 20  
Db 907 ATGATGCTGTGAAGAAACACAGAGGCCCGGGAGTCTCGAAGGTGAGAGGCCAGC 966  
QY 21 PROPROAlaValGInGInProthAspProAlaseProthValAlaThThProGlu 40  
Db 967 CCTCTGCTGTGCAGCAGCCACCTGACCCCGCATCTGCCATCTGCTACCAACGCTTGAG 1026  
QY 41 ProValGIsEAsPAlaGlyAsPlyAsnAlaThThValaGlyAspAsPglUProGlu 60  
Db 1027 CCCGTGGGCTCGATGCTGGGACAGAGATGCCCAAGCGAGCGATGACGAGCCAGAG 1086  
QY 61 TyrlGluaspGlyArGgGlyPheGlylleglyGluLeuValTrpGlyLysLeuArgGlyPhe 80  
Db 1087 TAGAGGACGGCCGGGCTTTGGCATTGGGAGCTGGTGTGGGAAACTCGGGGCTTC 1146  
QY 81 SerTrpTrpProGlyArGllValSerTrpTrpMetThrGlyArGserArGAla1aGlu 100  
Db 1147 TCCGTGTGCCAGGCCCATTTGTCTTGTGGATGAGTGAAGCGGCCGAGAGCGAGCTGAA 1206  
QY 101 GlyThrArGTrpValMetCTTPheGlyAspGlyLysPheSerValValCysValGluLys 120  
Db 1207 GGCACCCGCTGGTGCATGTGCTTGGAGACGGCAATTTCTAGTGTCTGTTGAGAG 1266  
QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThThYrAsnLysGlnPro 140  
Db 1267 CTGATGCGCGTGAAGCTGTTTTCAGTGGCTTCCACAGGCCAGTCAACAGACAGCCC 1326  
QY 141 MetYrArGlyAla1a1eYrGluValLeuGlnValAla1aSerSerArGAlaGlyLysLeu 160  
Db 1327 ATGTACCGCAAGCACTACAGAGTCTCGAGGTGCCAGACACCGCGGGAAGCTG 1386  
QY 161 PheProValCysHisAsPserAsPglUserAsPThrAlaLysAlaValaGluValGlnAsn 180  
Db 1387 TTCCTGGTGTCCACGACAGGATGAGATGACACTGCCAAGCCGTGGAGTGCAGAC 1446  
QY 181 LysProMetlleglUTPAlaLeuGlyGlyPheGlnProserGlyProLysGlyLeuGlu 200  
Db 1447 AAGCCATGATTGAATGAGGCCCTGGGGCTTCCAGCTTCTGGCCCTAAGGGCTGGAG 1506  
QY 201 ProProGluGluGluLysAsnProTyrlYsGluValYrThThAsPmetTrpValGluPro 220  
Db 1507 CCACAGAGAAAGAAAGAAATCCCTCAAGAAATGTACAGGACATGTGGTGAACCT 1566  
QY 221 GluAla1a1a1aYrAlaPProProProProAlaLysLysProArGlyLysSerThAlaGlu 240  
Db 1567 GAGCAAGCTGCTACGACACCACTCCACACCAAAAGCCCGGAGAGACAGCGAG 1626  
QY 241 LysProLysValLysGluIle1leAsPgluaRgThArGluarGleuValYrGluVal 260  
Db 1627 AAGCCCAAGGCAAGGAGATTATTGATGAGGCGCAAGAGAGCGCGTGTGTACAGGTG 1686  
QY 261 ArgGlnLysCysArgAsn1legluAsP1leCys1leSerCysGlySArLeuAsnValThr 280  
Db 1687 CGGAGAGAGTCCGGAACATTGAGAGCATTCCTCTGTGGAGCTCAAGTTAC 1746  
QY 281 LeuGluHisPProLeuPheValaGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
Db 1747 CTGGAACACCCCTCTGTTGGAGAGATGTGCCAAAAGTCAAGAACTGCTTCTGGAG 1806  
QY 301 CysAlaTyrlGlnYrAsPAsPAsPglYrGlnSerTyrlCysThrlleCysCysGlyGly 320  
Db 1807 TGTGCTACAGTACGACGAGACGAGCGGTACAGTCTTACATCTGCTGTGGGGC 1866

QY	321	ArgIuValLeuEmeCysGlyAsnAsnAnCysCysArgCysPheCysValGluCysVal	340
Db	1867	CGrGAGGTGCATGTGTGGGAAACAACTGTGTCCAGTGTCTTTTGGGTGAGATGTGTG	1926
QY	341	AspLeuLeuValGlyProGlyValAlaGlnAlaIleLysGluAsnProTrpAsnCys	360
Db	1927	GACCTTGTGGGGGCGGGGGGTGGCCAGCGACGCAATTAAGAAAGACCTTGAACGTG	1986
QY	361	TyrMetCysGlyHisLysGlyThrTyTGlyLeuLeuArgArgGluAsnTrpProSer	380
Db	1987	TACATGTGGGGGCACAAAGGTACTACCGGCTGTGGCGCGGAGAGATGTGGCCCTCC	2046
QY	381	ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr	400
Db	2047	CGGCTCCAGATGTTCTTGTGTAATTAACACAGACCGAATTTGACCTCCAAAGTTTAC	2106
QY	401	ProProValProAlaGluLysArgLysProIleArgValIleuSerLeuPheAspGlyIle	420
Db	2107	CCACCTGTCCAGGTGAGAAAGAGAACCCATCCGGGTGTGTCTCTTTGATGGATC	2166
QY	421	AlaThrGlyLeuLeuValIleuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer	440
Db	2167	GCTACAGGGCTCTCTGTGTCTAAGACTTGGGCATTCAAGTGGACCGGTACTATGGCTCG	2226
QY	441	GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr	460
Db	2227	GAGGTGTGTAGGACTTCATCAGCGTGGGATGGTGGCGGCACAGGGAGATCATGTAC	2286
QY	461	ValGluAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeu	480
Db	2287	GTCGGGGACGTCCCGACGCTCACACAGAAAGATATCCAGGAGTGGGGCCATTGATCTG	2346
QY	481	ValIleGlyLysSerProCysAsnAspPheSerIleValAsnProAlaArgLysGlyLeu	500
Db	2347	GTAATGGGGGCACTCCCTGCATAGACTTCTTCATGTCAACCTGTCTGGAGAGGCTC	2406
QY	501	TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro	520
Db	2407	TACAGAGGCACTGGCCGGCTCTTCTTGTAGTTTACCGCTCCGTCAATGACCGCGGCC	2466
QY	521	LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyVal	540
Db	2467	AAGAGGGAGATGATCGCCCTTCTTGTGGGCTTTGAAATGATGAGGCGCATGGGCGTT	2526
QY	541	SerAspLysAspAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys	560
Db	2527	AGTACAAAGAGGACATTCGCCGATTTCTCGAGTCCAACTGTGATGATTTGATGCCAA	2586
QY	561	GluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg	580
Db	2587	GAACTGTACGTGCACACAGGGCGCGTACTTTCGGGGTAACTTCCCGGTATGAACAGG	2646
QY	581	ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGlyCysLeuGluHisGlyArg	600
Db	2647	CCGTGTGCATCCACTGTGAATGAATGAAGCTGACGTGACGAGGTCTCTGAGCAAGCGAGG	2706
QY	601	IleLeuLysPheSerLysValArgThrIleThrThrArgSerSerAsnSerIleLysGlnGly	620
Db	2707	ATAGCCAAAGTTCACCAAAGTAGAGCACTTACTACAGAGGTCAAACTCCATTAACAGAGGC	2766
QY	621	LysAspGlnHisProProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGlu	640
Db	2767	AAAGACACAGCATTTTCTGTCTTATGAATGAAAGAGACATCTTATGTGTGACTGAA	2826
QY	641	MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu	660
Db	2827	ATGGAAGAAGGTATTTGTTTCCAGTCCACTATATCTGACGCTCCAACTACAGCGCGCTTG	2886
QY	661	AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla	680
Db	2887	GCGAGGCAGACATCTGTGGGCGCGGTCAAGAGCGGCGCAAGTCAATCCGCACATTGTGCT	2946

[illegible]

QY 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProGlyLeuGlu 200  
 Db 1548 AAGCCCATGATTGAATGGGCGCTGGGGGCTTCCAGCTTCTGGCCCTTAAGGGCGCTGGAG 1607  
 QY 201 ProProGluGluGluValAsnProGlyTyrGluValTyrThrAspMetTrpValGluPro 220  
 Db 1608 CCACCAAGAAAGAAAGAAATCCCTACAAAGAAAGTATACAGGACATGTGGGTGAACCT 1667  
 QY 221 GluAlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
 Db 1668 GAGGACACTGCTCTACGACCACTCCACACCAAAAAGCCCGGAAGAGCAGCGGAG 1727  
 QY 241 LysProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluVal 260  
 Db 1728 AAGCCCAAGGCTCAAGAGATTATGATGAGCCCAAGAGAGCGGCTGGTCTAGAGGTG 1787  
 QY 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 Db 1788 CGGCAAGAGTCCCGAATGAGACATCTGCATCTCTGTGGAGCCTCAATGTTACC 1847  
 QY 281 LeuGlnHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
 Db 1848 CTGGAACACCCCTCTTCTGTTGAGAGATGTGCCAAAACCTGCAGAACTGCTTCTGGAG 1907  
 QY 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
 Db 1908 TGTGCTGACCTACGATGACAGCAGCAGCGCTACCAAGTCTACTGCACCATCTGCTGGGGGC 1967  
 QY 321 ArgGluValIleuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
 Db 1968 CGTAGGGCTCATGTGCGGAAACAAACTGCTGACAGTGTCTTTTGGCTGAAGTGTG 2027  
 QY 341 AspleuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTrpAsnCys 360  
 Db 2028 GACTCTTGGTGGGGCGGGGGGCTGCCAGCAGCAGCATTTAAGAAACCTCGTGAATCGC 2087  
 QY 361 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgGluArgGluAspTrpProSer 380  
 Db 2088 TACATGTGCGGCGCAAGAGGTACCTACCGGCTGCTGCGCGCGAGGACTGCGCCCTCC 2147  
 QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400  
 Db 2148 CGGCTCCAGATGTTCTTCTGCTAATACACGACCGAATTGACCTCCAAAGGTTTAC 2207  
 QY 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
 Db 2208 CCACCTGTCCAGCTGAGAAAGAGGACCATCCGGGTGCTGCTCTCTTGATGCAATC 2267  
 QY 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
 Db 2268 GCTACAGGCTCTCTGCTGCTGTAAGCATTTGGGCATTCAGGTGAGCCGTTACCTCCG 2327  
 QY 441 GluValCysGluAspSerIleThrValGlyMetValAspHisGlnGlyLysIleuMetTyr 460  
 Db 2328 GAGGTGTGTGAGACTCATACGCTGGGACATGTGTCGGCACAGGAGGAATCATGTATC 2387  
 QY 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeu 480  
 Db 2388 GTCCGGGAGCGTCCGAGCGTCACACAGAGCATATCCAGAGTGGGCGCCATTCATCTG 2447  
 QY 481 ValIleGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db 2448 GTGATTTGGGGCAGTCTCTGCAATGACTCTTCATCTGCAACCTCTGCGCAGGCGCTC 2507  
 QY 501 TyrGluGlyThrGlyArgLeuPheGluPheGlyPheTyrArgLeuLeuHisAspAlaArgPro 520  
 Db 2508 TACAGGGCACTGGCGGCTCTTCTTAAGTTCTACCGCTCTGCAAGATGCGGGCGCC 2567  
 QY 521 LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyVal 540  
 Db 2568 AAGGAGGAGATGATGCGCCCTTCTTGTGGCTCTTGAAGAATGTGTGGCCATGCGCGCT 2627

QY 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
 Db 2628 AGTACAAAGGGGACATCTGCCGATTTCTGAGTCCAACTCTGTATGATTGATGCCAA 2687  
 QY 561 GluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg 580  
 Db 2688 GAAGTCTACCTGACACACAGGGCCCGCTACTTCTGGGGTAACTTCCCGATGAAACAGG 2747  
 QY 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGlyCysLeuGluHisGlyArg 600  
 Db 2748 CCGTTGGCATCTCACTGGAATGATTAAGCTGAGCTGCAGAGGTGTCTGGAGCATGGCAGG 2807  
 QY 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
 Db 2808 ATACCAAGTTCCACAAAGTGAAGACCATTACTACGAGTCCAACTCATTAAGCAGGGC 2867  
 QY 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGlu 640  
 Db 2868 AAGACCAAGCATTTCTCTGCTTCATGATGAGAAAGAGACATCTTATGTGTCACATGAA 2927  
 QY 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
 Db 2928 ATGCAAAAGGATATTGGTTTCCAGTCCACTATATGACGTCTCCAAATGAGCGGCTTG 2987  
 QY 661 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
 Db 2988 GCGAGGAGAGATCTGCTGGGCGGCTCATGAGCGTGCATCATCCGCACCTTTGGCT 3047  
 QY 681 ProLeuLysGluTyrPheAlaCysVal 689  
 Db 3048 CCGCTGAAGGAGTATTTGCGTGTGTG 3074  
 RESULT 9  
 LOCUS CQ720469 2841 bp DNA linear PAT 03-FEB-2004  
 DEFINITION Sequence 6403 from Patent WO2068579.  
 ACCESSION CQ720469  
 VERSION CQ720469.1 GI:42281326  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE  
 1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
 AUTHORS Knts, such as nucleic acid arrays, comprising a majority of  
 TITLE humanexons or transcripts, for detecting expression and other uses  
 thereof  
 JOURNAL Patent: WO 02068579-A 6403 06-SEP-2002;  
 FEATUES PE Corporation (NY) (US)  
 source Location/Qualifiers  
 1. 2841  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 2841  
 Score: 3746.00 Matches: 689  
 Percent Similarity: 98.0% Conservative: 0  
 Best Local Similarity: 98.0% Mismatches: 0  
 Query Match: 99.5% Indels: 14  
 DB: 2 Gaps: 1  
 US-10-623-813-86 (1-689) x CQ720469 (1-2841)  
 QY 1 MetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAlaSer 20  
 Db 730 ATGAACTCTGTGGAAGAAACCAAGGGCCCGGAGAGTCTCAAGAGGTGAGAGGCCAGC 789  
 QY 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40

Db 790 CCTCCTGCTGTGCAGCAGCAGCCACCTGACCCCGCATCCCGCATGTGGCTACCAACGCGCTTAC 849  
 Qy 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrValAlaGlyAspAspGluProGlu 60  
 Db 850 CCCCTGGGGTCCGATGTGCTGGGGACAAAGATGCCAACAGCAGGCGAGACGAGCAGAG 909  
 Qy 61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPhe 80  
 Db 910 TACAGAGACGCGCGGGCTTTGGCATTTGGGAGCTGTGTGGGGAACTCGGGGGCTTC 969  
 Qy 81 SerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
 Db 970 TCTCGTGGCCAGGCGCGCATTTGTGTGGATGACGGGCGGAGCCGAGCGAGCTGAA 1029  
 Qy 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys 120  
 Db 1030 GGCAACCGCTGGGTCACTGTGTGGAGACGCGCAAAATCTCAGTGTGTGTGAGAG 1089  
 Qy 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGluPro 140  
 Db 1090 CTGATGCGGCTGAGCTGTGTGTGAGTGCATGCCATCCAGCGCAGTACACAGCAGCGCC 1149  
 Qy 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 Db 1150 ATGTACCGCAAGCCATCTACGAGGTCTCTGCAAGTGGCCAGCAGCCGCGGGGAGCTG 1209  
 Qy 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsn 180  
 Db 1210 TTCCCGGTGTGCCAGCAGAGGATGAGTGCATCTGCCAAAGCGCTGTGAGTGCAGAAC 1269  
 Qy 181 LysProMetIleGluTyrPalaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
 Db 1270 AAGCCCATGTGATGATGGCCCTGGGGGCTTCCAGCCCTTGTGCGCTTAAGGGCTGAG 1329  
 Qy 201 ProProGluGluGlyLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro 220  
 Db 1330 CCACCGAAGAAGAGAAATCCCTACAAAGATGTACACGACATGTGGTGGAACT 1389  
 Qy 221 GluAlaAlaAlaTyrAlaProProProAlaLysLysProArgLysSerThrAlaGlu 240  
 Db 1390 GAGGACACTCTTCAGCACCACTCCACAGCCAAAGAGCCCGGAGAGACGCGGAG 1449  
 Qy 241 LysProLysValLysGluIleLeuAspGluArgThrArgGluArgLeuValTyrGluVal 260  
 Db 1450 AAGCCCAAGGCAAGGAGATATTGATGAGCGCAAGAGAGCGGCTGTGTAGAGGCTG 1509  
 Qy 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 Db 1510 CGGCAGAAAGTCCGGAACATTGAGACATCTGCATCTCTGTGGAGCCCTCAATGTACC 1569  
 Qy 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnAsnCysPheLeuGlu 300  
 Db 1570 CTGGAACACCCCTCTTCTGTGGAGATGTGCCAAACTGCAGAACTGCTTCTTGAG 1629  
 Qy 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
 Db 1630 TGTGCGTACAGTACGACGAGCAGCGCTACCAAGTCCATCTGCACTCTGTGGGGGC 1689  
 Qy 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
 Db 1690 CGGAGAGTGTCTATGTGCGGAAACACAACTGTGCGAGTGTCTTTGCGTGGAGTGTG 1749  
 Qy 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyrPasnCys 360  
 Db 1750 GACCTCTTGGTGGGCGGGGGCTGCCAGGACCAATTAAAGAAAGCCCTGGAATGTC 1809  
 Qy 361 TyrMetCysGlyHisLysGlyTyrTyrGlyLeuLeuArgArgArgGluAspTrpProSer 380  
 Db 1810 TATATGTGCGGGCAAGGGTACTTACGGGCTGTGTGGGGCGCAGAGGACTGGCCCTCC 1869  
 Qy 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLysValTyr 400

Db 1870 CGGCTCAGATGTTCTTGCTTAATAACCAAGCAGCAAGGATTTGACCTCCAAAGTTTAC 1929  
 Qy 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
 Db 1930 CCACCTGTCCAGGTGAGAAAGAGAAACCCATCGGGTGTGTCTCTCTTGTGATGAAATC 1989  
 Qy 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
 Db 1990 GCTACAGGCTCTCTGTGTGTAAGAGACTTGGGCAATTAGGTGAGCCGCTCATTTGCCCTC 2049  
 Qy 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
 Db 2050 GAGGTGTGAGGACTCTCATCAGGTGGCATGTGGCGCAGCAGGGAGATCATGTAC 2109  
 Qy 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrProPheAspLeu 480  
 Db 2110 GTGGGGACCTCCGACAGCGTACACAGAACATCCAGAGTGGGGCCCATTCGATCTG 2169  
 Qy 481 ValIleGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db 2170 GTGATGGGGGAGTCCCTGCAATGACCTTCCATGTCAACCTGTCGCAAGGCGCTC 2229  
 Qy 501 Tyr-----GluGlyThrGlyArg 506  
 Db 2230 TACGGTACAGCCCCAGCTGATGGCTTCTCTCCGACCTCTCAGAGGAGCATGGCCGG 2289  
 Qy 507 LeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArg 526  
 Db 2290 CTCTTCTTGTAGTTCACCCCTCTGCATGATGCGGGCCAAAGGGAGATGATGCG 2349  
 Qy 527 ProPhePheTrpLeuPheGluAsnValAlaAlaMetGlyValSerAspLysArgAspIle 546  
 Db 2350 CCTTCTTCTGCGCTTGTGAAGATGTGGTCCATGGCCGTGTGTGACAAAGGAGATC 2409  
 Qy 547 SerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHis 566  
 Db 2410 TCGGATTTCTCAGTCCAAACCTGTGATGATGATGCCAAAGAGTGCAGCTGCAC 2469  
 Qy 567 ArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrVal 586  
 Db 2470 AGGGCCCGCTACTTCTGGGGTAACTTCCGGTATGAACAGGCGGTGGCATCCACGTG 2529  
 Qy 587 AsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyValArgIleAlaLysPheSerLys 606  
 Db 2530 AATGATAGCTGAGCTGCAGAGTGTCTGAGCAATGCAAGATTAAGCAATTCAGTAA 2589  
 Qy 607 ValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHisPhePro 626  
 Db 2590 GTGAGGACATTACTACGAGGTCAAACTCCATTAAGCAGGCAAGGCCAGCATTTTCT 2649  
 Qy 627 ValPheMetAsnGluLysGluAspIleLeuTrpCysThrGluMetGluArgValPheGly 646  
 Db 2650 GTCTTCATGATGAGAAAGAGCATCTTATGTGTGACTGAAATGAGAAAGGATATTGGT 2709  
 Qy 647 PheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeu 666  
 Db 2710 TTCCAGTCCACTATACGTGCTGCTCAACATGAGCCGCTTGGCAGAGCAATGCTG 2769  
 Qy 667 GlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPhe 686  
 Db 2770 GCGCGGTCAATGAGCGTGCAGATCATCCGACCTTTCGCTCGGTGAAGGATATTTT 2829  
 Qy 687 AlaCysVal 689  
 Db 2830 GCCTGTGTG 2838  
 RESULT 10  
 AF480164 2318 bp mRNA linear ROD 14-OCT-2002  
 LOCUS AF480164  
 DEFINITION Mus musculus DNA cytosine methyltransferase 3A2 (Dnmt3a2) mRNA,  
 complete cds.  
 ACCESSION AF480164  
 VERSION AF480164.1 GI:23954439

KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Chen, T., Ueda, Y., Xie, S. and Li, E.  
TITLE A Novel Dmrtc1 Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Active de Novo Methylation  
JOURNAL J. Biol. Chem. 277 (41), 38746-38754 (2002)  
PUBMED 12138111  
REFERENCE 2 (bases 1 to 2318)  
AUTHORS Chen, T. and Li, E.  
TITLE Direct Submission  
JOURNAL Submitted (31-JAN-2002) Cardiovascular Research Center, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA

FEATURES  
source Location/Qualifiers  
1..2318  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="129S4/SvJae"  
/db\_xref="taxon:10090"  
/chromosome="12"  
/map="12A2-A3"  
1..2318  
/gene="Dmrtc1a2"  
148..2217  
/gene="Dmrtc1a2"  
/function="de novo DNA cytosine methylation"  
/codon\_start=1  
/product="DNA cytosine methyltransferase 3A2"  
/protein\_id="AA04038.1"  
/db\_xref="GI:23954440"  
translation="MNAVEENQSGESQKVEEAPPAQOPPTDASPPTVATTPPEVVG  
DAGKNATKADDEPEYEDGFGIGELVMKLGKGFSPWMPRIYSMWMTGSRAGECT  
RWVMFRGKRSVYCEVETKMLPSSRCSA.PHOATYKQMPYRKAIYEVQVASSAGKL  
FPACHDSDESQKAVYONKQIMBALGGOFSOPKLEPPEEKPIKEYITDMYV  
EPLEAAVAPPPAKKPRSTTEKPKVKIIDERTERLVYERQKRNIDICSCGS  
LNVLTLEHPLFIGMCKNCFLECAVYDDGYOSYCTICGGREVLKGNMNCRC  
FCVCGVDLVGGGAQAAIKEDPMNCWCGHGTGLRRREDMPSRLQMFANNDG  
EEDPKVYPPVPAEKRPRIVLSPDGIATGLVLDKLGIDVDRYIASEVEDSTTVG  
MYRQGIKIMYGVDSYTKIKIQEWGPELDYVIGSGPMDLSIVNAPRGGLYEGGRILF  
FEFRLHDAPKESGDRPFPMLENNVAVMGVSKRDI.SRPLESPWITDKESAAH  
RARYPWGNLPGMNRLASTVNDKLELOCLSHGRITAKSKVRTITTRNSIKQGDOR  
FPVPMNEKEDILMCTEMERVFVPHVYTDVSNMSRLARQRLIGRSMSVPIRLH.FAPL  
KEYFACV"

ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 2318  
Score: 3714.00 Matches: 679  
Percent Similarity: 98.84 Conservative: 2  
Best Local Similarity: 98.54 Mismatches: 8  
Query Match: 98.78 Indels: 0  
DB: 6 Gaps: 0  
US-10-623-813-86 (1-689) x AF480164 (1-2318)

QY 61 TyrgluaspGIyArGIyPheGlyIleGlyGluLeuValTrpGIyLysLeuArGIyPhe 80  
DB 328 TATAGAGATGGCCGGGCTTTGGCATTGGAGAGCTGGTGTGGGGAACTTCGGGGCTTC 387  
QY 81 SerTrpTrpProGIyArGIyleValSerTrpTrpMetThrGIyArGSerArGIAlaGlu 100  
DB 388 TCCGTGTGGCCAGGCCCAATGTGTCTTGGGTGATGACAGGCCGAGCCGACGCTGAA 447  
QY 101 GlyIhrTrpTrpValMetTrpPheGIyAspGIyLysPheSerValValCysValGluLys 120  
DB 448 GGCACTGGCTGGGTGATGTGTGGAGATGGCAAGTTCATGATGTGTGTGGAGANG 507  
QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyraAsnLysGlnPro 140  
DB 508 CTCATGGCGCTGAGACTCTCTTCTGACATGCATTCACACAGGCCACCTACACAGAGGCC 567  
QY 141 MetTrpArGIyAlaIleTyrgluValLeuGlnValAlaSerSerArGIAlaGlyLysLeu 160  
DB 568 ATGTACCGCAAGCCATCTACGAAGTCTCTCCAGCTGGCCAGACCCGTCGGGAACTG 627  
QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValAsn 180  
DB 628 TTTCACGCTTGCATGACATGATGAAGTACAGTGGCAAGCTGTGGAACTGCAGAAC 687  
QY 181 LysProMetIleGIyTrpAlaLeuGlyGIyPheGlnProSerGIyProLysGIyLeuGlu 200  
DB 688 AAGCAGATGATTGAATGGGCCCTCGTGGCTTCCAGCCCTCGGGTCTTAAGGCCCTGAG 747  
QY 201 ProProGluGluGluLysAsnProTyrgluValTyrgluValLysProArGIySerThrAlaGlu 220  
DB 748 CCACCAAGAGAGAGAAATCCCTTACAGGAATTTACCCGACATGTGGTGGAGCT 807  
QY 221 GluAlaAlaAlaTyraIleProProProAlaLysLysProArGIySerThrAlaGlu 240  
DB 808 GAACAGCTGCTTACGGCCCAACCCCAACCAAGAAACCCAGAAAGACACACAGAG 867  
QY 241 LysProLysValLysGluIleIleAspGluArGIyThrArGIyArGIyVal 260  
DB 868 AACCTTAAGGCACAGAGATCATTTAGAGGCAACAGAGCGCTGTATAGAGGTG 927  
QY 261 ArgGluLysCysArGIanIleGluAspIleCysIleSerCysGIySerLeuAsnValThr 280  
DB 928 CGCCAGAGGTGCACAAAATCTCGAAGCAATTTGTTCATGTGGAGCTTAAGTCAACC 987  
QY 281 LeuGluHisProLeuPheValGIyLysCysGlnAsnCysLysAsnCysPheLeuGlu 300  
DB 988 CTGAGAGCACTCCATCTTATTTGAGGCAATGTGCCAGAACTTAAGAACTCTTCTTGAG 1047  
QY 301 CysAlaTyrgluTrpAspAspGIyTyrgluSerTyrcysThrIleCysCysGIyGIy 320  
DB 1048 TGTGCTTACCAAGTATGACAGATGGGTAACAGTCTATTCACACATCTGCTGTGGGGG 1107  
QY 321 ArgGluValLeuMetCysGIyAsnAsnAsnCysCysArGIyPheCysValGluCysVal 340  
DB 1108 CGTGAAGGCTCAATGTGTGGGACAAACATCGTCGACGGTCTTTGTGTGAGTGTGTG 1167  
QY 341 AspLeuLeuValGIyProGIyAlaAlaGlnAlaIleLysGIyAspProTyraAsnCys 360  
DB 1168 GATCTCTTGGGGGCAAGAGCTGCTCAGGACCACTTAAGAAAGACCTTGGAATCTGC 1227  
QY 361 TyrMetCysGIyHisLysGIyThrTyrgluLeuLeuArGIyAspTrpProSer 380  
DB 1228 TACATGTGGCGGACATAAGGCACTTATGGGCTGTGGAAAGACGGAACTGGCTTCT 1287  
QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLysValTyrglu 400  
DB 1288 CGACTCCAGATGTTCTTTGCAATACCATGACCAAGATTTGACCCCAAGGTTTAC 1347  
QY 401 ProProValProAlaGluLysArGIyLeuProIleArGIyValLeuSerLeuPheAspGIyIle 420  
DB 1348 CCACCTGTGGCACTGTGAGAGAGAACCCATCCGCGTCTGTCTCTTTGATGGAGATT 1407  
QY 421 AlaThrGluLeuLeuValLeuLysAspLeuGlyIleGlnValAspArGIyTrpIleAsn 440

Db 1408 GCTACAGGCTCTCGTGGCTGAAGAGCACTGGGCACTCCAGTGAGCCGCTACATTCCTCC 1467  
 Qy 441 GtUva1Cysgluaspser11ethrVal1GlymetVal1ArgH1sg1ng1Lyse1MetTyr 460  
 Db 1468 GAGGTGTGTAGGAGCTCCATCACCGTGGGCGATGTGTGGCGGACACGGGAAAGATCATGTAC 1527  
 Qy 461 Val1GlyaspVal1AngserVal1ThGlnLysH1s1leg1ng1UTPp1GlyPropHaspLeu 480  
 Db 1528 GTGGGGAGCTCCGACGCTACACAAAGCATATCCAGAGGTGGGCGCCATTTGACCTG 1587  
 Qy 481 Val1leg1LyseProCysAsnAspLeuSer11Val1AsnProAlaArgLyse1Leu 500  
 Db 1588 GTGATTGAGGAGCACTCCCTCCAAAGCACTCTCCATTGTCAACCTGCCCGGAAGGACTT 1647  
 Qy 501 TyrGluGlyThrGlyArgLeuPheGluPheGlyTyrArgLeuH1sAspAlaArgPro 520  
 Db 1648 TATGAGGGTACTGGCCGCTCTTCTTGTAGATTCTAACCGCTCTGCATGTAGCCGCGCC 1707  
 Qy 521 LysGluGlyAspAspArgProPhePheTyrLeuPheGluAsnVal1Val1AlaMetGlyVal 540  
 Db 1708 AAGGAGGAGATGATGCCCTCTTCTGTGCTTTAGAAATGTGTGGCCATGGGCGGTT 1767  
 Qy 541 SerAspLysArgAsp11SerArgPheLeuGluSerAsnProVal1Met11LeAspAlaLys 560  
 Db 1768 AGTGACAGAGGAGCATCTGCCGATTTCTTGAGCTTAACCCCGTGATGATGACGCCAA 1827  
 Qy 561 GluValSer11Ala1ArgVal1ArgTyrPheTyrP1LysAsnLeuProGlyMetAsnArg 580  
 Db 1828 GAGGTGTCTCTGCACACAGAGGCCGCTTACTTGTGGGTAACTTCTGCGCATGACAGG 1887  
 Qy 581 ProLeu1AserThrVal1AsnAspLyseGluLeuGlnGlyCysLeuGluH1sGlyArg 600  
 Db 1888 CTTTGGCATCCACTGTGAATGATACCTGAGCTGCAAGAGTGTCTGGAGCAGGAGAGA 1947  
 Qy 601 L1Ala1LysPheSerLysVal1ArgThr11eThrThrArgSerAsnSer11eLysGlnGly 620  
 Db 1948 ATAGGCCAAGTTCAGCAAGAGAGGACCATTAACACAGGTCAACTTAATAAGCAGGCG 2007  
 Qy 621 LysAspGlnHisPheProVal1PheMetAsnGluLysGluAsp11eLeuTyrCysThrGlu 640  
 Db 2008 AAGAGCAGCATTTCCCGCTTCTCATGAACGAGAAAGAGGACATCTGTGTGACACTGAA 2067  
 Qy 641 MetGluArgVal1PheGlyPheProVal1HisTyrThrAspVal1SerAsnMetSerArgLeu 660  
 Db 2068 ATGGAAGGGTGTGGCTTCCCGCTCACACAGACGCTTCAACATGAGCGGCTTG 2127  
 Qy 661 AlaArgGlnArgLeuGlyArgSerTyrSerVal1ProVal11eArgH1sLeuPheAla 680  
 Db 2128 GCGAGGCGAGACTGCTGGGCGGCACTGTGAGCGTGCCTGATCCGCCACCTTCTGCT 2187  
 Qy 681 ProLeuLyseGluTyrPheAlaCysVal1 689  
 Db 2188 CCGCTGAAGAAATATTTTGTGTGTG 2214

RESULT 11  
 LOCUS BC007466 4094 bp mRNA linear ROD 29-JUN-2004  
 DEFINITION Mus musculus DNA methyltransferase 3A, transcript variant 1, mRNA  
 (cdna clone MGC:5662 IMAGE:3492853), complete cds.  
 ACCESSION BC007466  
 VERSION BC007466.1 GI:13938620  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 4094)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Dege, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,

Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Shaplenko, M., Soares, M.B., Donald, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Udén, T.B., Teshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahy, J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, J.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Snailus, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 4094)  
 Strausberg, R.  
 Direct Submission  
 Submitted (01-MAY-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.sngc.stanford.edu>  
 Contact: (Dickson, Mark) [mcdpaxil.stanford.edu](mailto:mcdpaxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>  
 Series: IRAC Plate: 8 Row: n Column: 11  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6681208.  
 Location/Qualifiers  
 1..4094  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Czech 11"  
 /db\_xref="taxon:10090"  
 /clone="MGC:5662 IMAGE:3492853"  
 /tissue\_type="Mammary tumor metastasized to lung. Tumor  
 arose spontaneously from a senescent normal mammary  
 (clonal) outgrowth infected with the virus MMTV."  
 /clone\_lib="NCI CGAP Lu29"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 1..4094  
 /gene="Dnmt3a"  
 /db\_xref="GeneID:13435"  
 /db\_xref="MGI:1261827"  
 245..2971  
 /gene="Dnmt3a"  
 /codon\_start=1  
 /product="DNA methyltransferase 3A, isoform 1"  
 /protein\_id="AAH07466.1"  
 /db\_xref="GI:13938621"  
 /db\_xref="GeneID:13435"  
 /db\_xref="MGI:1261827"  
 /translation="MPSSGPGDTRSSSLEREDDRKGEBOENRKGKEROEPATARK  
 VRRPKRKRHPYESSDTRKDPATVTKKSCQMADSDPSDLRPGDLKESRSEPOEES  
 PAAQGGGGAABSEGTETPEASRAVNGCCTVKEGRGASAGCKEKPTNISMKE  
 GSGRGRGGLGWSLSLRQRPMLPTFOADPYITSKRKDEWLARWKRLEKAKAVIA  
 VNMAVEENOASGESQKVEASPPAVOOPDPAFTVAATPEVGGAGDKNATKAAD



QY 621 lyeaspglnhisrheprovalphmetasnlglyuaspilieleutrcythrclu 640  
 DB 2762 AAGACGACGATTTCCCGCTTTCATGACGAGGAGGACATCTGTGTGACTGAA 2821  
 QY 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
 DB 2822 ATGAAAGGGGCTTTGGCTTCCCGCTCCACTACACAGCGCTCCACATGAGCGCTTG 2861  
 QY 661 AlaArgGlnArgLeuGlyArgSerTspSerValProValIleArgHisLeuPheAla 680  
 DB 2882 GCGAGCAGAGACTGCTGGCGGCGATCGGAGCGCTGCGGCTCATCCGCACTCTCGCT 2941  
 QY 681 ProLeuArgGlyTyrPheAlaCysVal 689  
 DB 2942 CCGCTGAAGAAATATTTGCTTGCTG 2968  
 RESULT 12  
 AF068625 4192 bp mRNA linear ROD 06-DEC-1999  
 LOCUS Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF068625 GI:6449467  
 VERSION AF068625  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 4192)  
 AUTHORS Okano, M., Xie, S., and Li, E.  
 TITLE Cloning and characterization of a family of novel mammalian DNA  
 (cytosine-5) methyltransferases  
 JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
 PUBMED 9662389  
 REFERENCE 2 (bases 1 to 4192)  
 AUTHORS Xie, S., Okano, M., and Li, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 REFERENCE 3 (bases 1 to 4192)  
 AUTHORS Okano, M., Chijiwa, T., Sasaki, H., and Li, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 REMARK Sequence update by submitter  
 COMMENT On Nov 18, 1999 this sequence version replaced gi:3327977.  
 FEATURES  
 source  
 1. 4192  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /chromosome="12"  
 /map="4.0 cM"  
 1. 4192  
 /gene="Dnmt3a"  
 /gene="Dnmt3a"  
 1. 4192  
 /note="similar to EST sequences deposited in GenBank  
 Accession Numbers AA052791, AA111043, AA154890, AA240794,  
 AA756653, W58898, W59299, W91664, and W91665"  
 217. 2943  
 /gene="Dnmt3a"  
 /function="de novo DNA methylation"  
 /note="contains Cys-rich region"  
 /codon\_start=1  
 /product="DNA cytosine-5 methyltransferase 3A"  
 /protein\_id="AAC0177.2"  
 /db\_xref="GI:6449468"  
 /translation="MPSGPGDTSSSSLEREDRKEGEDEBNKNGKEEROPSPATARK  
 VGRGRKRHPVSSDTPKDPATVTKSCMAQDPSDLPLNGDLKRSSEPOPEES  
 PAAGKGGAPAEAGSGTETPEASRAVENGCCVTKEGAGAGGCKGKQINISMKE

GSRGRRLGGWESSLRORPMPLTFOAGDPVYISKRKDEMLARWKEAEKAKVIA  
 VMAVAENMASSGSOKVEASPPAVOOPDPAETVATTEPVGADGDNATKAAD  
 EPEYKMPGSGISELIVMGKLRGSSWMPGRVSMHMTGRSAAGTRVMFGCKRSV  
 VCVEKMLPSSFCSAFHQATNTKQPMYRKAIYEVLOASSRAGLPACHDSBDSG  
 KAVEVONKOMIEMALGFGFQSPGRKLEPPEENKFPKEYVTDMVVEEAALAPPRA  
 KKRKSTTEKPKYKEIIDERTRERLYVEYOKRNIEDICISGSLVTEHPLEFIG  
 MCQNKQFLECAYOYDDGOSYCTICCGKRVLMGNNCCRCFCEVCVDLLVGG  
 AACAAIKEDPMNVMCGHGGTGLLRREBMPRLQMPFANNDOEPDPRKVPVPA  
 EKRKPRVSLSPFGIATGLVLKDLGIQUDRYIASVCDSITVGMRHQKITVVD  
 VRSVTKHIOEMKPFDLVIGSSPCNDLSIVNPKGLYBGTGLPFFFTLHDARK  
 EGDDRPFMLFENVNVMGVSDKRDISRFESNVPMTDAKESAAHRAHYWGLPGNN  
 RPLASTVNDLELQECLEHRIKFSKVRTITRRSNSIKGKXOHPFVPMKEEDILW  
 CTEMERVFGRPVHYTVDVSNMSRLARQRLGRSWSVPVIRHLFAPLKEYFACV"

## ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 4192  
 Score: 3714.00 Matches: 679  
 Percent Similarity: 98.8% Conservative: 2  
 Best Local Similarity: 98.5% Mismatches: 8  
 Query Match: 98.7% Indels: 0  
 DB: Gaps: 0

US-10-623-813-86 (1-689) x AF068625 (1-4192)

QY 1 MetAsnAlaValGlnGluAsnGlnGlyProGlyGluSerGlnTyrValGlnGluAlaSer 20  
 DB 874 ATAAATGCTGTGAAAGAAACCAAGCCTCTGAGAGCTGAGAGGTGAGAGGAGCCAGC 933  
 QY 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40  
 DB 934 CTCCTGCTGTGACGACGACCCAGGACCTGCTTCCGACTGTGGCCACACCCCTGAG 993  
 QY 41 ProValGlySerAspAlaGlyAspIlyAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
 DB 994 CCAAGTGAAGGAGATCTGGGAGCAAGAAATGCTACCAAGACCCGACGTGAGCCAGC 1053  
 QY 61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPhe 80  
 DB 1054 TATGAGATGCGCGGGGCTTTGGCATTTGGAGACTGGTGGGGGAAACTTCGGGGCTTC 1113  
 QY 81 SerTyrTrpProGlyArgIleValSerTyrTrpMetThrGlyArgSerArgAlaAlaGlu 100  
 DB 1114 TCCTGGTGGCCAGCCGGAATTGTGTCTTGGTGATGACAGCGGAGCCGAGAGCTGAA 1173  
 QY 101 GlyThrArgTyrValMetTyrPheGlyAspGlyLysPheSerValValCysValGluLys 120  
 DB 1174 GGCACCTGCTGGGCTCATGTGTTGCGAGATGGCAAGTTCTCACTGTGTGTGAGAGAG 1233  
 QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140  
 DB 1234 CTATGCGCGTGAGCTCTCTCTCAAGTCATTCACAGCCAGCCCTTCACACAGACACCC 1293  
 QY 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 DB 1294 ATGTACCGAAGAACCATCTACAGAACTCTCAAGTGGCCAGCGCGTGGGAAAGCTG 1353  
 QY 161 PheProValCysHisAspSerAspGluSerThrAlaLysAlaValGluValGlnAsn 180  
 DB 1354 TTTCGAGCTTGCATACAGTGTGAAAGGACAGTGGCAAGCTGTGGAAAGTGCACAAAC 1413  
 QY 181 LysProMetCileGluTyrPheGlnGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
 DB 1414 AACCAAGTATTAAGGGGCTCGGTGGCTTCCAGCCCTCGGCTCTTAAGGGCTGGAG 1473  
 QY 201 ProProGluGlnGluLysAsnProTyrLysGluValTyrThrAspMetTyrValGluPro 220  
 DB 1474 CCACCAAGAAAGAAAGAAATCTTCAAGGAAGTTTACCCGACATGTGGTGGAGCCT 1533  
 QY 221 GluAlaAlaAlaTyrAlaProProProProAlaLysLysArgArgLysSerThrAlaGlu 240  
 DB 1534 GAAGCAGCTGCTTACCCGCCACCCACGACCCAAAGAAACCCAGAAAGAACACAGAG 1593

QY 241 LysProluValIySgUuIleIleaspGluArGrThraGrGluArGrLeuValIyTrGluVal 260  
 DB 1594 AAACCTAAGGTCAGAGAGATCATTCATAGCGCAAGAGCGCGCTGGTGTATGAGGTG 1653  
 QY 261 ArgGlnIySArGAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 DB 1654 CGCAGAGAGTCAGAAACATCGAGAGCATTTGTATCTCATGTGGAGCGCTCAATGTCAAC 1713  
 QY 281 LeuGluHisProluPheValGlyGlyMetCysGlnAsnCysGlySArGpHeCysVal 300  
 DB 1714 CTGAGACCCCACTTCTTATGAGGCAATGTGCCAAGACTGTAGAACTGCTTCTTGAG 1773  
 QY 301 CysAlaTrGrInTrArSpAspAspGlyTrGrInserTrCysThrIleCysCysGly 320  
 DB 1774 TGTCTTACCGATAGACGATGGTGTACCACTCTTTCACCATCTGCTGTGGGGG 1833  
 QY 321 ArgGluValLeuMetCysGlyAsnAsnCysCysArGpHeCysValGluCysVal 340  
 DB 1834 CGTGAAGTCTCATGTGTGGAAACAACTGCTCAAGTCTTTGTGTGAGTGTG 1893  
 QY 341 AspLeuLeuValIGlyProGlyValAlaIleAlaIleValIySgUuAspProTrpAsn 360  
 DB 1894 GATCTCTTGTGGGGCCAGAGCTGTCTCAGCAGCATTTAAGAAAGACCCCTGGAATGC 1953  
 QY 361 TyrMetCysGlyIyHsIySgIyThrTrGlyLeuLeuArGrArGrGluAspTrpProSer 380  
 DB 1954 TACATGTGGGGCATTAAGGCACTATGGCTGTGCCAAGACGGGAAGACTGGCTTCT 2013  
 QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisaspGlnIyPheAspProCysValIyTr 400  
 DB 2014 CGACTCCAGATGTTCTTTGCCAATACCATGACACGAAATTTGACCCCCCAAGCTTAC 2073  
 QY 401 ProProlValProAlaGluValArGlySerProIleArGValLeuSerLeuPheAspGly 420  
 DB 2074 CCACCTGTGCCAGCTGAGAAAGAAAGGCCCAATCGCGTGTCTCTTTGATGGGATT 2133  
 QY 421 AlaThrGlyLeuLeuValLeuIySgUuIleGlnValAspArgTrIleAlaSer 440  
 DB 2134 GCTACAGGGCTCCTGTGCTCTAAGGACCTGGGCATCCAAAGTGGACCGTACATTCCTCC 2193  
 QY 441 GluValCysGluAspSerIleThrValIyMetValArGhIeGlnGlyIySgIyMetTr 460  
 DB 2194 GAGGTGTGTGAGCATCATCATCGGTGGCATGTGTGGCCGACCAAGAAATCATCTATAC 2253  
 QY 461 ValGlyAspValArGserValThrGlnIyHsIleGlnGlyIyTrGlyProPheAspLeu 480  
 DB 2254 GTCGGGAGCTCCGACAGGCTCACAGAAAGATATCCAGAGTGGGGCCCATTCGACTG 2313  
 QY 481 ValIleGlySerProCysAsnAspLeuSerIleValIleAspProAlaArgIySgIyLeu 500  
 DB 2314 GTGATTGGAGGCACTCCCTGCATGACTCTTCATGTTCMAACCTGCCCCGCAAGGACTT 2373  
 QY 501 TyrGlnGlyThrGlyArGrLeuPhePheGluPheTrArGrLeuLeuHisaspAlaArgPro 520  
 DB 2374 TATAGAGGTAATGAGCCGCTCTTCTTGAATTACCCCTCTGATGATGCGGGGCC 2433  
 QY 521 LysGlnGlyAspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyVal 540  
 DB 2434 AAGAGGAGATGATGCGCCCTTCTTGGGCTTTTGAAATGTGGTGGCCATGGGCTT 2493  
 QY 541 SerAspIySArGAspIleSerArGrPheLeuGluSerAsnProValMetIleAspAlaIyS 560  
 DB 2494 AGTACAAAGAGGACATCTCCGATTTCTTGAAGTAAACCCCGATGATGATGACGCCAA 2553  
 QY 561 GluValSerAlaAlaHisArGAlaArgTrPheTrpGlyAsnLeuProGlyMetAsnArg 580  
 DB 2554 GAATGTCTCTGCACACAGGGCCGCTTACTTGGGGTAACTTCCCTGGCATGAAACGG 2613  
 QY 581 ProLeuAlaSerTrpValAsnAspGlySgUuLeuGlnGlyCysLeuGlnHisGlyArG 600  
 DB 2614 CTTTGGATCCACTGTGAATGATAGCTGGAGCTGCAGAGAGTGTCTGGACACGGCGGA 2673  
 QY 601 IleAlaIySPhSerIySValArGrThrIleThrTrArGserAsnSerIleIySgInGly 620

DB 2674 ATACCAAGTTCAGCAAGTGAAGGACCATTAACCCAGTCAACTCTATAAGCAGGGC 2733  
 QY 621 LysAspGlnHisPheProValPheMetAsnGluIySgUuAspIleLeuTrpCysThrGlu 640  
 DB 2734 AAAAGCAGACATTTCCCGCTTCTATAGACAGAAAGAGACATCTGTGTGACGTGA 2793  
 QY 641 MetGluArgValPheGlyPheProValHisTrYrThrAspValSerAsnMetSerArgLeu 660  
 DB 2794 ATGAAAGGGTGTGGCTTCCCGTCCATACACAGCGTCTCCAAACATAGAGCGCTTG 2853  
 QY 661 AlaArgGlnArGrLeuGlyArGrSerTrpSerValProValIleArGhIleLeuPheAla 680  
 DB 2854 GCGAGCAGAGACTGTGGCCGATCGTGAAGCGTGGCGGTATCCGCCACTTGTGCT 2913  
 QY 681 ProLeuIySgUuTrPheAlaCysVal 689  
 DB 2914 CCGCTGAAGAAATATTTGCTGTGTG 2940  
 RESULT 13  
 LOCUS AY271299 2597 bp mRNA linear MAM 05-SEP-2003  
 DEFINITION Bos taurus DNA methyltransferase 3a isoform 4 (DNMT3A) mRNA,  
 partial cds.  
 ACCESSION AY271299  
 VERSION AY271299.1 GI:32350982  
 KEYWORDS  
 SOURCE Bos taurus (cattle)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
 Pecora; Bovidae; Bovinae; Bos.  
 1 (bases 1 to 2597)  
 Golding,M.C. and Westhusin,M.E.  
 Analysis of DNA (cytosine 5) Methyltransferase mRNA Sequence and  
 Expression in Bovine Preimplantation Embryos, Fetal and Adult  
 Tissues  
 JOURNAL Gene Expr. Patterns 3 (5), 551-558 (2003)  
 REFERENCE 2 (bases 1 to 2597)  
 AUTHORS Golding,M.C. and Westhusin,M.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-APR-2003) Veterinary Physiology, Texas A&M - CVM, 500  
 University Dr. West, College Station, TX 77843-4466, USA  
 FEATURES  
 source location/Qualifiers  
 1..2597  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 <1..2597  
 /gene="DNMT3A"  
 <1..2467  
 /gene="DNMT3A"  
 /note="DNMT3A4; DNA (cytosine 5) methyltransferase"  
 /codon\_start=2  
 /product="DNA methyltransferase 3a isoform 4"  
 /protein\_id="AAP75902.1"  
 /db\_xref="GI:32350983"  
 /translation="EGEEOEARAKEEROBOSTTARKVGRGRRKXHPVSSSDTPD  
 PATYTSLSAADSGBSELTYPNDLEKRSPOEBEGSPAGOCAGABEGBAETEPD  
 EASRAVENGCCTREPLOKAEKRAKATIAVNAVEBQGSTGQKEASBPAAVOQT  
 DPASPVAATTPREVGADAGDKNTKAADPEYEDGFGIIGELVWKLGFSPWPKR  
 IVEWMTGSRRAAEGTRVMMFGDKFSVVCVSKLMPLSFCSAFHQAATYKOPMYK  
 AIEVLCOVASRAAGKLPVWCHDDESDTAAVEVONKOMIEMALGCPGSPGKLEBP  
 EEEKRYKRYVTMMVPEPAAYAAPPAKPKRSTKEPKVEIIDERTRRLVVEY  
 ROKCRINEDICTSGSLNTLLEHPLFIGMCCQKCNQLECAYOYDDGYOYCTTCC  
 GGRBVLMCGNNNCCRCFCVBCVDLLVGGAAQAIKBDPMNYCYCHGKTYGLRRD  
 DWSRRLQMPFAANNHDEFDPPKVPVPAPAKRPIRVLSLFDGATGLVKDGLIY  
 DRYIASEVCEDSITVGAVRHQKIMYGVDSVYOKAIQWGPDLVIGSPCNDLSI  
 VNPAKGLYEGRGLFEFFYRLIHDARPKBGDRPFMWLEENVVLGVSDKDISRL  
 ENSPVMIDAKEVLAARBARYPWGNLPPQMNPLASTVUDKLELOECLSHGRIAKSKYR  
 TITTRNSITGCGDQHFPPVPMNEKEDILWCTEMERVGFVHTVDVSNMRLLARQLL  
 GRSMSPVIRHLFAPLKEYPACV"

ORIGIN

Alignment Scores:			
Pred. No.:	0	Length:	2597
Score:	3712.00	Matches:	678
Percent Similarity:	99.1%	Conservative:	5
Best Local Similarity:	98.4%	Mismatches:	6
Query Match:	98.6%	Indels:	0
DB:	14	Gaps:	0
US-10-623-813-86 (1-689) x AY271299 (1-2597)			
QY	1	MetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGluValGluGluAlaSer	20
DB	398	ATAAATGCTCTTAAAGAAACCAAGGCTCCACGAGCTCAGAAAGTGGAGAGCCAGT	457
QY	21	ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrProGlu	40
DB	458	CTCCCGCTGTGACAGCCACCGACCTGCGTCCCAACAGTGGCCACCGCCTGAG	517
QY	41	ProValGlySerAspAlaGlyAspIysAsnAlaThrIlysaIaGlyAspAspGluProGlu	60
DB	518	CCCGTGGGGCCCATGCGGGGACAAAGATGCCAACAGCAGCTACATGAACCGGAG	577
QY	61	TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTrrpGlyIlySLeuArgIlyPhe	80
DB	578	TACGAGGACGGCCGGGCTTTGGCATTGGGGAGCTGGTGGGGGAAACTCGGGGCTTC	637
QY	81	SetTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaIaGlu	100
DB	638	TTCGTGTGGCCAGCCCGCATTTGTGTGTGTGATGACGGGCGGAGCCGAGCCGGA	697
QY	101	GlyThrArgTrpValMetTrpPheGlyAspGlyIlySPheserValValCysValGluIlyS	120
DB	698	GGCAGCTGTGGGTCAATGGGTTCGAGACGGCAAGTTCAGTGGTGTGGAGAG	757
QY	121	LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnIlySGlnPro	140
DB	758	CTGATCCCGCTGAGCTCTCTTCAGTCTTTCCACAGGCCACCTACAAACAGCAGCCC	817
QY	141	MetTyrArgIlysaIaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyIlySLeu	160
DB	818	ATGTACCGCAAGGCTATCTACGAGGTCTCTCAGGTGGCCAGCATCGAGGGGGAAGCTG	877
QY	161	PheProValCysHisAspSerAspGluSerAspThrAlaIysAlaValGluAlaAsn	180
DB	878	TTTCCAAATGTCATGACGACGAGAGGACACCTGCCAAGCCGTGGAGGTGCAGAAC	937
QY	181	LysProMetIleGluTrpAlaLeuGlyIlyPheGlnProSerGlyProIlySGlyLeuGlu	200
DB	938	AAACAAATGATCAAGTGGGCTTGGAGGTTCCAGCCCTCTGGCCCCCAAGGGCTGGAG	997
QY	201	ProProGluGluGluIlySAspProTyrIlySGluValTyrThrAspMetTrpValGluPro	220
DB	998	CCCCGAGAGAGGAGAAACCCCTTACAAAGATTACACAGACATGTGGTGAACCC	1057
QY	221	GluAlaAlaAlaIlyrAlaProProProProAlaIlySlySProArgIlySerThrAlaGlu	240
DB	1058	GAGGAGCTGTGCTATGGCGGCCCCCAAGCCAAAAGCCCGGAAAGACACACGAG	1117
QY	241	LysProIlyValIlySGluIleIleAspGluArgThrArgIlyuArgLeuValTyrGluVal	260
DB	1118	AAACCTTAAGTCAAGAGATCATTTGAAGACGCAAGAGAGGCGGTGTGTAGAGTA	1177
QY	261	ArgGlnIlySerCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr	280
DB	1178	CGGCAAGAGTGGCGGAACATCGAGACATCTGCTCTTGTGGAGCCCTCAACGTACC	1237
QY	281	LeuGluHisProLeuPheValIlyGlyIlyMetCysGlnAsnCysIlySAsnCysPheLeuGlu	300
DB	1238	TTGGAAACACCTCTTCTTCAATCGAGAGATGTGCCAAAACGCAAGAACTCTCTTGGAA	1297
QY	301	CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyIly	320

DB	1298	TGGCGTACAGTACGATGAGTATCAGTCTTCACTGACCACTCTGCGGGGCG	1357
QY	321	ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal	340
DB	1358	CGGAGAGTGTCAATGTGTGGAAACAATGTGTGAGGGTCTTTTGGCGGAGATGTGTG	1417
QY	341	AspLeuLeuValGlyProGlyAlaIaGlnAlaIaIleIlySGluAspProTyrAsnCys	360
DB	1418	GATCTCTTGTGGGGCCAGGGGCCGCGAGCGCCATCAAAAGAACCCCTGGAACTGC	1477
QY	361	TyrMetCysGlyIlySlySGlyThrTyrGlyLeuLeuArgArgIlyuArgIlyuArgTrpProSer	380
DB	1478	TACATGTGGCGGACAAAGGCACTTACGGGCTGCTCGGCGCGGAGACGACTGGCGCTCT	1537
QY	381	ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProIlySValTyr	400
DB	1538	CGGCTCAGATGTTCTTCCCAACAACATGACAGAGAAATTCATCTCCGAAGTTTAC	1597
QY	401	ProProValProAlaGluIlySArgIlySProIleArgValLeuSerLeuPheAspGlyIle	420
DB	1598	CCACCTGTCCAGCTAGAGAGAGAAAGCCATCCGGGTGCTGTCTATTCGATGGAATT	1657
QY	421	AlaThrGlyLeuLeuValLeuIlySAspLeuGlyIleGlnValIlySArgTyrIleAlaSer	440
DB	1658	GCTACAGGGCTTGTGGTGTGAAGAGACTTGGGCAATTCAGGTGACCGCTACATCGCTCC	1717
QY	441	GluValCysGluAspSerIleThrValIlyMetValArgHisGlnGlyIlySIlleMetTyr	460
DB	1718	GAGGTGTGAGGACATCCATCACGTGGGATGTGCGGACCAAGGAAAGATCATATAC	1777
QY	461	ValGlyAspValArgSerValThrGlnIlySHisIleGlnIlyTrpGlyProPheAspLeu	480
DB	1778	GTCGGGACGTCCGACGCTTACACAGAACATATCCAGAAATGGGCGCGTTCGATCTG	1837
QY	481	ValIleGlyIlySerProCysAsnAspLeuSerIleValAsnProAlaArgIlySGlyLeu	500
DB	1838	GTGATTTGGGGGAGTCCCTGCAATGATCTCTCATGTCTCAACCTCGCCCAAGGGCTC	1897
QY	501	TyrGluGlyThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAspAlaArgPro	520
DB	1898	TACGAGGACATGGCGGCTCTTGTAGTTTACCGCTCTTGCAATGATGGCGGCCC	1957
QY	521	LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValAlaIleMetGlyVal	540
DB	1958	AAAGAGGAGATACCGGCCCTCTTCTGCTCTTGAAGATGTGGTGGTCCGGGGCTT	2017
QY	541	SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaIlyS	560
DB	2018	AGTGACAAAGGGACATCTCGCATTTCTCGAGTCCAAACCTGTGATGATGATGCCAAA	2077
QY	561	GluValSerValAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg	580
DB	2078	GAGGTCTACGTGGGACAGGGCCGCTACTTCTGGGGGAACTTCTGGTATGAACAGG	2137
QY	581	ProLeuAlaSerThrValAsnAspLysLeuGluGlnIlyCysLeuGlnHisGlyArg	600
DB	2138	CCATTGGCATCCACTGTGAATGATTAAGCTGAGCTCAGAGAGTGTCTGAGACAGCGCGA	2197
QY	601	IleAlaLysPheSerIlySValArgThrIleThrThrArgSerAsnSerIleIlySGlnGly	620
DB	2198	ATAGCCAAAGTTCCAGAAAGAGGACCATTAATACTAGGTGGAATCCATTAAGCAGGGC	2257
QY	621	LysAspGlnHisPheProValIlePheMetAsnGluIlyGluAspIleLeuTrpCysThrGlu	640
DB	2258	AAAGACACAGATTTCCTCCGTTTCATGAATGAAGAAAGAGACATTTATGTGCACTGAA	2317
QY	641	MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu	660
DB	2318	ATGGAAGGGTGTTTGGCTTCCCTGTCCACTATACGAGAGTCTCCAACATGAGCCGGTTG	2377
QY	661	AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla	680
DB	2378	GCGAGCAGAGACTGTGGGCGGTGTGTGAGCGTCCGGTCAATCGGCCACTTTCGCT	2437

QY 681 ProLeuLySGluTYrPheAlaCysVal 689  
 Db 2438 CCGCTGAAGAAATATTGCTGTGCTG 2464  
 RESULT 14  
 AY271298  
 LOCUS AY271298 2798 bp mRNA linear NAM 05-SEP-2003  
 DEFINITION Bos taurus DNA methyltransferase 3a (DNMT3A) mRNA, partial cds.  
 ACCESSION AY271298  
 VERSION AY271298.1 GI:32350980  
 KEYWORDS  
 SOURCE Bos taurus (cattle)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
 Pecora; Bovidae; Bovinae; Bos.  
 1 (bases 1 to 2798)  
 Golding, M.C. and Westhusin, M.E.  
 Analysis of DNA (cytosine 5) Methyltransferase mRNA Sequence and  
 Expression in Bovine Preimplantation Embryos, Fetal and Adult  
 Tissues  
 Gene Expr. Patterns 3 (5), 551-558 (2003)  
 2 (bases 1 to 2798)  
 Golding, M.C. and Westhusin, M.E.  
 Direct Submission  
 Submitted (08-APR-2003) Veterinary Physiology, Texas A&M - CVM, 500  
 University Dr. West, College Station, TX 77843-4466, USA  
 FEATURES  
 source  
 1..2798  
 Location/Qualifiers  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 <1..2798  
 /gene="DNMT3A"  
 <1..2668  
 /gene="DNMT3A"  
 /note="DNMT3A; DNA (cytosine 5) methyltransferase"  
 /codon\_start=2  
 /product="DNA methyltransferase 3a"  
 /protein\_id="AAP75901.1"  
 /db\_xref="GI:32350981"  
 /translation="MGEQEEBAKAEKEQEBSTTARKVGRGRKKRPVSSDPKQ  
 PAVTSKLSMAQDSGSELYPNGLDEKSESPQEPESGAGCKGKAPRGGATETPR  
 EASAVENGCCCTPKDGRGAPAEKGEQETNIEMKMGSGRLRGGLGWESLQR  
 MPRLTFOAGDPYVYSKRKDEMLARKWEAKKAKVLAVMVAENQSTSQVEEA  
 SPFAVOPTDPASPTVATTPPVAGADGKNAKTAADPEPEDEGFGIGELVWGL  
 RGFEWMPGRIVSMWMTGRSRAEGRTWMDGDFSVCEKLMPLSSFCSAFHQAT  
 YNKQPMYRKAIYEVLQVASSRAGLTPMKCHSDSDPAKAYENKMKIMIALGROP  
 SGFGLPEPEEKRPYKSVYTDNWEPEALAIAPPPAKRKSTTEPKYKEITDER  
 TRELVYEVRRQCRNIEDI CISCGLNLTLEHPLFTIGMCQNCNCFLECAVYDDG  
 YQSVCTICCGREVLMCGNNCCRCFCVCEVDLVPGLAALAIKEDPMWYMGHG  
 TYGLRRDDWPSRLQMFANNDQEPFPKYRPPVPEKKAPIRVLSLFGIATGL  
 VLKQLGVDWRYIASVECEDSITVGMVHOKIMVGVRSVTOHGIQEMFPDLVIG  
 GSPNDLSIVNPAKILYEGTGRLPFEEYRLLHDARPEGGDRPFFMIFENVVLGVS  
 DKRISIRLESNPVMTDAKEVSAHARAYFNGNLPKMRPLASTVNDLLEIQECLHS  
 RIAPFSKRITITTSNSIKGKDHPFVFNMEKEDILMCTEMERVFGEVPHYTVDSNM  
 SRLRORLLGRSWSVPVIRHLFAPLKEYFACV"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 2798  
 Score: 3712.00 Matches: 678  
 Percent Similarity: 99.1% Conservative: 5  
 Best Local Similarity: 98.4% Mismatches: 6  
 Query Match: 98.6% Indels: 0  
 DB: 14 Gaps: 0  
 US-10-623-813-86 (1-689) x AY271298 (1-2798)  
 QY 1 MetAsnAlaValAGluGluAsnGlnGlyProGlyGluSerGlnLysValAGluGluAlaSer 20  
 Db 599 ATGATGCTGTTGAAGAAACAGAGGCTCAACCGAGTCTCAGAAAGTGAAGAGGCCAGT 658

QY 21 ProProAlaValAGlnGlnProThrAspProAlaSerProThrValAlaThrThProGlu 40  
 Db 659 CCTCCCTGTCGACGACCCACCGACCTGCGTCCGCCACAGAGGCGCACCCCTGAG 718  
 QY 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
 Db 719 CCCGTGGGCGCGCATGCCCGGGACAAATGCCCAAGCACTGACGTGAACCCGAG 778  
 QY 61 TyGluAspGlyArgGlyPheGlyLeGlyGluLeuValTrpGlyLysLeuArgGlyPhe 80  
 Db 779 TACGAGACGGCCGGGCTTTGGCATTGGGAGCTGTGTGGGGAAACTCGCGGGCTTC 838  
 QY 81 SerTrpTrpProGlyArgGlyValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
 Db 839 TCCGTGGGCGGACCGCCCATTTGTCTTGTGTGAATGACGGGCGGACCGAGCGGAA 898  
 QY 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValAlaCysValGluLys 120  
 Db 899 GGCACCTGTTGGTCATGTGTTGCGAGACGGCAAGTTCTCAGTGTGTGTGTGAGAG 958  
 QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyAsnLysGlnPro 140  
 Db 959 CTGATGCCGCTGAGCTCTTCTGCAGTGCTTCCACACGAGCCACTCAACACAGCCGCC 1018  
 QY 141 MetTyArgLysAlaLeuTyArgLysValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 Db 1019 ATGTACCGAAGCTATCTACAGAGTCTCTGAGGTGCGACGACGTCAAGCGGGAGCTG 1078  
 QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValAGluGlnAsn 180  
 Db 1079 TTTTCAATGTGCCATGACAGAGAGAGACGACCTCCAAAGCGCGTGAAGTCAAGAC 1138  
 QY 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
 Db 1139 AAACAATGATCGAGTGGGCTGTGGAGGTTCCAGCCCTCTGCGCCCAAGGCTTGAG 1198  
 QY 201 ProProGluGluGluLysAsnProTrpLysGluValTyThrAspMetTrpValGluPro 220  
 Db 1199 CCCCAGAGAGGAGAAACCCCTCAAGAAAGTTTACACAGACATGTGGTTGAACCC 1258  
 QY 221 GluAlaAlaAlaTyAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
 Db 1259 GAGGACACTGCTATGCGCCGCCCCCAACAGCCAAAGCCGGAAGAGCAACGAG 1318  
 QY 241 LysProLysValLysGluIleIleAspGluArgThrArgLysAlaGluValTyGluVal 260  
 Db 1319 AAGCTAAGGTCAAGAGATCATTTGATGAACGCAACAAGAGCGGCTGTGTACAGGTA 1378  
 QY 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerIleAsnValThr 280  
 Db 1379 CGGAGAGAGTCCGGAACATCGAAGACATCTGCTCTTGTGGAGCTTCAAGTCAAC 1438  
 QY 281 LeuGluHisProLeuPheValAGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
 Db 1439 TTGGAACACCTCTCTTCATCGAGGAATGTCCAAAACTGCAGAAAGACTGCTCGAA 1498  
 QY 301 CysAlaTyArgLysTyAspAspAspGlyTyGlnSerTyCysThrIleCysCysGlyGly 320  
 Db 1499 TGCCTGACCAAGTACGATGACGATGAGCTATCATCTACATGCAACATCTGCTGGGGGG 1558  
 QY 321 ArgGluValLeuMetCysGlyLysAsnAsnCysCysArgCysPheCysValAGluCysVal 340  
 Db 1559 CGCAGGTGCTCATGTGTGGAAACAAATTGCTGCGAGTGCTTTTTCGTGGATGTGTG 1618  
 QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTrpAsnCys 360  
 Db 1619 GATCTCTTGGTGGGCGCAGGGGCGCGCAGGACCATCAAGAAGACCCCTGGAATGCG 1678  
 QY 361 TyMetCysGlyHisLysGlyThrTyGlyLeuLeuArgArgArgGluAspTrpProSer 380  
 Db 1679 TACATGTGCGGCGCAAGGGGCACTTACGAGGCTGCTGCGGGCGGCGGACGACTGCGCT 1738

[illegible]

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1
AUTHORS	Totoki,Y., Toyoda,A., Takeda,T., Sakaki,Y., Tanaka,A., Yokoyama,S., Ohara,O., Nagase,T. and F.Kikuno,R.
TITLE	None Title
JOURNAL	Published Only in Database (2005)
REFERENCE	2 (bases 1 to 4476)
AUTHORS	Totoki,Y., Toyoda,A., Takeda,T., Sakaki,Y., Tanaka,A., Yokoyama,S., Ohara,O., Nagase,T. and F.Kikuno,R.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAR-2005) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research, 2-6-7 Kazusa-kametari, Kisarazu, Chiba, 292-0818, Japan [E-mail:cdminto@kazusa.or.jp, URL:http://protein.gsc.riken.go.jp/, Tel:81-438-52-3930, Fax:81-438-52-3931]
COMMENT	This work was supported in part by the National Project on Protein Structural and Functional Analysis, Ministry of Education, Culture, Sports, Science and Technology of Japan. Totoki Y, Toyoda A, Takeda T, Sakaki Y, Tanaka A, Yokoyama S. RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan. e-mail: aktanaka@postman.riken.go.jp URL: http://protein.gsc.riken.go.jp/.
FEATURES	location/Qualifiers
source	1..4476
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="pJ01914"
	/tissue_type="brain"
	/note="This clone is also named as hsk003000130 vector:pbInescriptII SK plus"
gene	<1..4476
	/gene="DNA cytosine methyltransferase 3 alpha isoform a variant"
CDS	<148..2583
	/gene="DNA cytosine methyltransferase 3 alpha isoform a variant"
	/inference="non-experimental evidence, no additional details recorded"
	/note="Start codon is not identified."
	/codon_start=1
	/product="DNA cytosine methyltransferase 3 alpha isoform a variant"
	/protein_id="BAD92070.1"
	/db_xref="GI:62087246"
	/translation="ISPKPTTITISPSPSSPPAPRSPSLPGAGOMPAMPSSGPDITSSSAEREDBKRDGEDEEPKRGKROPTTARKVGRGRKRKPVESGPTPDPAVITSSPSNAQDSGASGLPNDGLERSEPOPEGSVAGQKGAFAEGGAETLEPARAVENGCCPTREGGAPAEAGKEQKETNISMKMGSGRGLRGGLGWSSLRORWPLPLTPAGGPYITISKRKDEMLARKVEAEAKAKAVIAGNAVVENGGPESQVEASPPALVOOPTPASPVTATTPPEPGSDAGDKNATKAGDDEPEYEDRGFGIGELVWGLKRGFSWMDRIVSWMMGCRSRAAEGTRLVMMFGGKSEVVCVEKLMPLSPCSAPHQATYNKQPMYRKATILEVLOYASSRAKGLPVCVHDSBESDPAKAVENONKMITENALGQFQSPKXGLEPPEEKNPYKEVYTDMMVEEAIAVAPPPAKPRKSTAKPKVKEIIDEETRRLVYEVEOKRNIETDICIISGLNVTILEHPFVFGMQNCNCELECYQYDDGQYOSYCTICCGGSERLQMGNNKCRFCFCEVCDLVLAGQAQAIKEBPNNCVMGHGKTYGLLRRRESERLQMGNNKCRFCFCEVCDLVLAGQAQAIKEBPNNCVMGHGKTYGLLQGVDRYIASVCEYDSTIVGMWRHOGKTIYGVDSVYQXKHQEMKPPDLVIGGSPCLNDISTVNPARKGLYEGTGRLLFEFYRLLDHARKEGDGDFPFLFENVANMGVSDKDDISRFLEVNQRSHLY"
ORIGIN	
Alignment Scores:	
Pred. No.:	0
Score:	3702.00
Percent Similarity:	91.1%
Best Local Similarity:	91.1%
Query Match:	98.4%
	5
	length: 4476
	Matches: 689
	Conservative: 0
	Mismatches: 67
	Gaps: 1

US-10-623-813-86 (1-689) x AB208833 (1-4476)

QY 1 MetAsna1aValGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAsnSer 20

Db 907 ATGAATGCTGTGAGAGAAAACAGAGGGCCCGGGAGTCTCAGAAAGTGGAGAGGCCAGC 966

QY 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40

Db 967 CCTCTGTGTGTGACAGCCCACTACCCCGCATCCCTCAGTGTGCTACCAAGCCTGAG 1026

QY 41 ProValGlySerAspAlaGlyAspLysAsna1aThrLysAlaGlyAspAspGluProGlu 60

Db 1027 CCCGGGGGTCCGAGTCTGGGGAACAAGATGCCACCAAGCAGGGGATGACAGCCAGAG 1086

QY 61 TyrGluAspGlyArgGlyPheGlyTllegLysLysLeuValTyrGlyLysLeuArgGlyPhe 80

Db 1087 TACGAGGACGGCGGGGCTTTGGCATTTGGGAGCTGTGTGGGGAAACTCGCGGGCTTC 1146

QY 81 SerTPTTPProGlyArgLLeValSerTPTTPMetThrGlyArgSerArgLAlaGlu 100

Db 1147 TCCGTGTGCGCAGCGCGCATGTGTCTTGTGTGATGACGGCGCGGAGCCGAGCAGCTGA 1206

QY 101 GlyThrArgTTPValMetTTPPheGlyAspGlyLysPheSerValValCysValGluLys 120

Db 1207 GGCACCCCTGGGTATGT 1266

QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrThrAsnLysGlnPro 140

Db 1267 CTGATGCCGCTGAGCTCGTTTTCAGTGGCTTCCACAGGCCACGCTACCAACAGAGCC 1326

QY 141 MetTyrArgLysAlaLLeTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160

Db 1327 ATGTACCCCAAGCATCTACAGAGTCTTCAGAGTGGCCAGCAGCGCGCGGGAGAGCTG 1386

QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsn 180

Db 1387 TTCCTGGTGTGCGACGACGAGTGAAGTGCATCTGCCAAGCCGTGAGGTGCGAGAAC 1446

QY 181 LysProMetLlegLUTTPAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200

Db 1447 AAGCCATGATTGAATGGGCTCGGGGGCTTCCAGCCTTCTGAGCCCTTAAGGGCTGAG 1506

QY 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTTPAlaGluPro 220

Db 1507 CCACCAAGAGAGAGAAAGATCTCTACAAAGAGTGTACACGAGCATGTGGGTGAACCT 1566

QY 221 GluAlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240

Db 1567 GAGGACGTGCTTACGACACACTCCACAGCCAAAAGCCCGGAGAGGACAGCGGAG 1626

QY 241 LysProLysValLysGluLLeLLeAspGluArgThrArgGluArgLeuValTyrGluVal 260

Db 1627 AAGCCCAAGGTCAAGAGATTTATGTAGAGCCACAGAGAGAGCGGCTGTATCAAGGTG 1686

QY 261 ArgGlnLysCysArgAsnLlegLysPheLysLysSerCysGlySerLeuAsnValThr 280

Db 1687 CGCGAAGAGTGCAGAACATTGAGGACATCTGCATCTCTGTGGAGCCTCATGTATAC 1746

QY 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300

Db 1747 CTGGAACACCCCTCTCTGTGTGAGGAATGTCCAAAATGCAAGACGCTTTCTGGAG 1806

QY 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrLLeCysCysGlyGly 320

Db 1807 TGTGTGTACAGTACAGACACAGCGCTACAGTCTTACTGACACATCTGCTGTGGGGC 1866

QY 321 ArgGluValLLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340

Db 1867 CGTGTGTGTGTATG 1926

QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaLLeLysGluAspProTTPAsnCys 360

Db 1927 GACCTTGT 1986

QY 361 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgArgGluAspTTPProSer 380

Db 1987 TACATGT 2046

QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400

Db 2047 CGGCTCCAGATGTTCTTGT 2106

QY 401 ProProValProAlaGluLysArgLysProLLeArgValLLeuSerLeuPheAspGlyLLe 420

Db 2107 CCACCTGTCCCAAGCTGAGAGAGAGAGCCCATCCGGGTGTGTGTGTGTGTGTGTGTGT 2166

QY 421 AlaThrGlyLeuLeuValLLeuLysAspLeuGlyTllegValAlaAspArgTyrLLeLLeSer 440

Db 2167 GCTACAGGGCTCTGT 2226

QY 441 GluValCysGluAspSerLLeThrValGlyMetValArgHisGlnGlyLysLLeMetTyr 460

Db 2227 GAGGT 2286

QY 461 ValGlyAspValArgSerValThrGlnLysHisLLeGlnGluTTPGlyProPheAspLeu 480

Db 2287 GTCCGGGACGTCCCGCAGGCTCACAGAGCATATCCAGAGTGGGGCCATTCGATCTG 2346

QY 481 ValLLeGlyLysProCysAsnAspLeuSerLLeValAsnProAlaArgLysGlyLeu 500

Db 2347 GTGATGTGGGGAGTCCCTGTGAAATGACCTTCTTCATCGTCAACCTGCTCGCAGAGCC 2406

QY 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLysHisAspAlaArgPro 520

Db 2407 TACAGAGCACAATGCGCGCTTCTTGTGAGTTCTACCGCTCTGTGATGTGGGGGCC 2466

QY 521 LysGluGlyAspAspArgProPhePheTTPLeuPheGluAsnValValAlaMetGlyVal 540

Db 2467 AAGAGGGAGATGTATGCGCCCTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2526

QY 541 SerAspLysArgAspLLeSerArgPheLeu----- 560

Db 2527 AGTGCAAGAGGACATCTCGCATTTCTTCAGAGTCAAAAATAGCCACTGTACTAGTGG 2586

QY 550 ----- 570

Db 2587 GCTAGTGGCACTGAATGAACAGTGTAGAGCAACCCACCATCGCAGAGTTCATTTC 2646

QY 550 ----- 570

Db 2647 GACAATGCTGATCTAGAACGTTCCAGTGCAGCTTGTGAAATGTGGCTAGGTGAATTCC 2706

QY 551 -----GluSerAsnPr 554

Db 2707 AGCTTCTCTGAGAGCTGACAGCTAACCCAGTGTGTGCTCTGAGAGAAATCCAAACC 2766

QY 554 OValMetLLeAspAlaLysGluValLLeSerAlaAlaHisArgLLeArgTyrPheTTPGlyAs 574

Db 2767 TGTGATGATTTATGTGCAAAAGAGTGTGACAGACAGGGCCCGCTACTTCTGGGGTAA 2826

QY 574 LLeuProGlyMetAsnArgProLeuAlaSerThrValAlaAsnAspLysLeuGluGlnGly 594

Db 2827 CTTTCCCGGTATGAACAGCGCGTGTGACATCTGTGATATATGAAGTGTGAGAGTGCAGA 2886

QY 594 UCyLeuGluHisGlyArgLLeAlaLysPheSerLysValArgThrLLeThrThrArgSe 614

Db 2887 GTGTGTGAGCATGTGAGAGATGACCAAGTTCAGCAAAATGAGACATTACTACAGAGTCT 2946

QY 614 LAsnSerLLeLysGlnGlyLysAspGlnHisPheProValPheMetAsnGluLysGluAs 634

Db 2947 AAATCTCATTAAGCAGGGCAAAAGACACAGCATTTCTCTCTTCAATGAATGAGAAAGAGA 3006

QY 634 PLeuLeuTTPCysThrGlyMetGluArgValPheGlyPheProValHisTyrThrAspVal 654

Db 3007 CATCTTATGT 3066

OY	654	1SerPsmMetSerArgLeuAlaArgGlnArgLeuGlnArgSerTrpSerValProVa	674
DB	3067	CTCCACATGAGCCGCTTGCGAGGCAAGAGACTGCTGGGCGGTCATGAGGCTGCCAGT	3126
OY	674	111eArGHisLeuPheAlaProLeuLysGlnTyrPheAlaCysVal	689
DB	3127	CATCGGCCACCTCTTCGCTCGCTGAGGAGATTTTGGGTGTGTG	3172

Search completed: October 5, 2006, 21:09:00  
 Job time : 10579 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 5, 2006, 21:00:34 ; Search time 2088 Seconds  
(without alignments)  
6082.026 Million cell updates/sec

Title: US-10-623-813-86

Perfect score: 3763

Sequence: 1 MNVENVQSGESQKVEAS.....MSVPVIRHLPAPLKEYFACV 689

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xjh  
-O=/abs/ABSSWB.spool/US10623813/runat\_05102006\_121153\_27814/app\_query.fasta\_1  
-DB=Published Applcations NA Main -OEMT=fastcap -SUFFIX=p2n.rnphm  
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1  
-MATRIX=biosum62 -TRANS=human40.cdt -LIST=45 -DOCCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext  
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02h  
-USER=US10623813 @CGN 1.1 1675 @runat\_05102006\_121153\_27814 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBIOCK=100 -LONGLOG -DEV TIMEOUT=120  
-MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applcations NA Main:

1: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US08\_PUBCOMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
11: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
12: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
13: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
14: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US11B\_PUBCOMB.seq:\*  
15: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US11C\_PUBCOMB.seq:\*  
16: /EMC\_Celerra\_SIDS3/pcodata/2/pubpna/US11D\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3763	100.0	2371	9	US-10-623-813-84

2	3763	100.0	3005	6	US-10-144-577-2	Sequence 2, Appl1
3	3763	100.0	3005	7	US-10-172-118-132	Sequence 132, App
4	3763	100.0	3005	8	US-10-342-887-132	Sequence 132, App
5	3763	100.0	3009	11	US-10-330-773-279	Sequence 279, App
6	3763	100.0	4293	5	US-09-720-086-3	Sequence 3, Appl1
7	3763	100.0	4293	9	US-10-623-813-3	Sequence 3, Appl1
8	3714	98.7	2318	9	US-10-623-813-83	Sequence 83, Appl
9	3714	98.7	4192	5	US-09-720-086-1	Sequence 1, Appl1
10	3714	98.7	4192	9	US-10-623-813-1	Sequence 1, Appl1
11	3714	98.7	6157	11	US-10-330-773-276	Sequence 276, App
12	2713	72.1	2938	7	US-10-264-237-953	Sequence 953, App
13	2169.5	57.7	4195	5	US-09-720-086-2	Sequence 2, Appl1
14	2169.5	57.7	4195	9	US-10-623-813-2	Sequence 2, Appl1
15	2157	57.3	4145	5	US-09-720-086-4	Sequence 4, Appl1
16	2157	57.3	4145	6	US-10-144-577-3	Sequence 3, Appl1
17	2157	57.3	4145	6	US-10-144-577-38	Sequence 38, Appl
18	2157	57.3	4145	7	US-10-172-118-1337	Sequence 1337, Ap
19	2157	57.3	4145	8	US-10-342-887-1337	Sequence 1337, Ap
20	2157	57.3	4145	9	US-10-623-813-4	Sequence 4, Appl1
21	1878.5	49.9	3897	6	US-10-144-577-39	Sequence 39, Appl
22	1793.5	47.7	95484	11	US-10-330-773-275	Sequence 275, App
23	1771.5	47.1	127722	11	US-10-330-773-278	Sequence 278, App
24	1318	35.0	1060	6	US-10-106-698-926	Sequence 926, App
25	1095.5	29.1	709	3	US-09-969-034-1235	Sequence 1235, Ap
26	1003	26.7	711	10	US-10-450-763-21969	Sequence 21969, A
27	949	25.2	632	3	US-09-764-891-272	Sequence 272, App
28	942	25.0	994	12	US-10-301-480-60665	Sequence 60665, Sequence 122074,
29	942	25.0	994	12	US-10-301-480-122074	Sequence 122074,
30	872.5	23.2	2057	6	US-10-144-577-40	Sequence 40, Appl
31	785	20.9	2008	6	US-10-144-577-41	Sequence 41, Appl
32	750	19.9	440	5	US-09-720-086-54	Sequence 54, Appl
33	750	19.9	440	9	US-10-623-813-54	Sequence 54, Appl
34	693	18.4	2951	10	US-10-450-763-21970	Sequence 21970, A
35	689	18.3	1705	8	US-10-479-934-35	Sequence 35, Appl
36	645.5	17.2	1670	8	US-10-479-934-21	Sequence 21, Appl
37	640	17.0	1521	5	US-10-479-934-7	Sequence 7, Appl1
38	636	16.9	522	5	US-09-720-086-61	Sequence 61, Appl
39	636	16.9	522	9	US-10-623-813-61	Sequence 61, Appl
40	629	16.7	588	6	US-10-027-632-199022	Sequence 199022,
41	629	16.7	588	6	US-10-027-632-199023	Sequence 199023,
42	629	16.7	588	7	US-10-027-632-199022	Sequence 199022,
43	629	16.7	588	7	US-10-027-632-199023	Sequence 199023,
44	609	16.2	435	5	US-09-720-086-18	Sequence 18, Appl
45	609	16.2	435	9	US-10-623-813-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-10-623-813-84  
Sequence 84, Application US/10623813  
Publication No. US20040234997A1  
GENERAL INFORMATION:  
APPLICANT: Li, En  
APPLICANT: Okano, Masaki  
APPLICANT: Xie, Shaping  
TITLE OF INVENTION: De Novo DNA Cytosine Methyltransferase Genes, Polypeptides & Uses  
FILE REFERENCE: 0609.456003  
CURRENT APPLICATION NUMBER: US/10/623, 813  
CURRENT FILING DATE: 2003-07-22  
PRIOR APPLICATION NUMBER: US 09/720, 086  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: PCT/US99/14373  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: US 60/090, 906  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: US 60/093, 993  
PRIOR FILING DATE: 1998-07-24  
NUMBER OF SEQ ID NOS: 119  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 84

```

:   LENGTH: 2371
:   TYPE: DNA
:   ORGANISM: Homo sapiens
US-10-623-813-84

Alignment Scores:
Pred. No.:      0      Length:      2371
              3763.00  Matches:      689
Score:          100.0%  Conservative: 0
Percent Similarity: 100.0%  Mismatches: 0
Best Local Similarity: 100.0%  Indels: 0
Query Match:      9      Gaps:      0

US-10-623-813-86 (1-689) x US-10-623-813-84 (1-2371)

Oy      1  MetnsnAlaValGlUGluAsnGlnGlyProGlyGlyUserGlnUysValGlUGluLaser  20
Db      217  ATGAATGCTGTGGAGAAACCAAGGGGCCCGGGAGTCTCAGAAAGGTGAGAGGCCAGC  276

Oy      21  ProProAlaValAGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu  40
Db      277  CTTCTGCTGTGGACAGACCCCACTACCCCGCATCCCACTGTGGTACCAAGCCTTGAAG  336

Oy      41  ProValGlySerAspAlaGlyAspUysAsnAlaThrUysAlaGlyAspAspGluProGlu  60
Db      337  CCGGTGGGGTCCGATGCTGGGGACAAAGATGCCACCAAGCAGCGGATGACGAGCCAGAG  396

Oy      61  TyrGluAspGlyA-gGlyPheGlyY1IeGlyUleuVal1TTPGlyUysLeuArgGlyPhe  80
Db      397  TACAGAGCAGCGCCGGGCTTTGGCATTTGGGAGCTGTGTGGGGAACTCGGGGGCTTC  456

Oy      81  SerTrpTrpProGlyArgGlyLeuValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu  100
Db      457  TCCGTGTGGCCAGGGCCGCAATGTGTGTGTGGATGAGAGGGCCGAGGCCAGCACTGAA  516

Oy      101  GlyThrArgTrpValMetTrpPheGlyAspGlyUysPheSerValValCysValGluUys  120
Db      517  GGCAACCGCTGGGTGATGTGGTGGAGACGGCAAAATCTCAGTGTGTGTGTGAGAAG  576

Oy      121  LeuMetProLeuSerSerPheCysSerAlaPheHISGlnAlaThrThrAsnUysGlnPro  140
Db      577  CTGATGCCGCTGAGCTCTTTTTCAGTGCCTTCCACAGGCCCACTAACCAAGCAGGCC  636

Oy      141  MetTyrArgUysAlaIleTyrGlyUalLeuGlnValAlaSerSerArgAlaGlyUysLeu  160
Db      637  ATGATCCCAAGAGCCATTTACAGAGTCTTGCAGGTGGCCAGCAGCGCGGGAGAGCTG  696

Oy      161  PheProValCysHISAspSerAspGluSerAspThrAlaUysAlaValGluValAsn  180
Db      697  TTCCCGGTGTGCCACGACAGAGCATAGAGTACATGGCAAGCGCGGTGAGGTGCAGAAC  756

Oy      181  LysProMet1IeGlyTTPAlaLeuGlyGlyPheGlnProSerGlyProUysGlyLeuGlu  200
Db      757  AAGCCATGATTTGAATGGCCCTTGGGGGCTTCCAGCCTTCTGGGCCCTTAAGGGCTTGAG  816

Oy      201  ProProGluGluGluUysAsnProTyrUysGlyUal1TyrThrAspMetTrpValGluPro  220
Db      817  CCACCAAGAAAGAGAAAGATCTCTACAAAGAGTTCACGAGACATGTGGGTGAACTT  876

Oy      221  GluAlaAlaAla1AlyrAlaProProProProAlaUysUysProArgUysSerThrAlaGlu  240
Db      877  GAGGAGAGTGGCTTACGACACACTCCACAGCCAAAGAGCCCGGAAAGCAGCAGAG  936

Oy      241  LysProUysValUysGluU1IeUysPglUatGThrArgGluArgLeuVal1TyrGluVal  260
Db      937  AAGCCCAAGGTCAGAGGATTAATGATAGCGCACAAAGAGCGGCTGTGTACAGAGTG  996

Oy      261  ArgGlnUysCysArgAsn1IeGluAsp1IeCys1IeSerCysGlySerLeuAsnVal1Thr  280
Db      997  CGGCAAGAGTCCCGAATTCAGACATCTGCACTCTCTGTGGAGACCTCAATTTAC  1056

Oy      281  LeuGluHISProLeuPheValGlyUysMetCysGlnAsnCysUysAsnCysPheLeuGlu  300

```

```

Db      1057  CTGGAAACACCCCTCTTGTTGGAGATGTGCCAAACTGCAGAACTGCTTTCGAG  1116
Oy      301  CysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThr1IeCysCysGlyUys  320
Db      1117  TGTGCGTACCAATGACAGAGAGAGCGGTACCAAGTCTTACACACATCTGTTGGGGGC  1176

Oy      321  ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGlyCysVal  340
Db      1177  CGTGAGGTGCTCATGTGCGGAAACACACATGCTGCAAGGTCTTTTGTGGAGTGTG  1236

Oy      341  AspLeuLeuValGlyProGlyValAlaGlnAlaAla1IeUysGlyUasProTTPAsnCys  360
Db      1237  GACCTTGTGGGGCCGGGGGCTCCAGGACCACTTAAGAAAGACCCCTGAAATGTC  1296

Oy      361  TyrMetCysGlyHISUysGlyThrTyrGlyUleuLeuArgArgGlyUasPTTProSer  380
Db      1297  TCAATGTGGGCAAGAGGTACTTACGGGTCTGTGGCGCGCAGAGAGACTGGCCCTCC  1356

Oy      381  ArgLeuGlnMetPhePheAlaAsnAsnHISAspGlnGluPheAspProProUysVal1Tyr  400
Db      1357  CGGCTCCAGATGTTCTTGGTAATTAACAGACAGGAAATTTGACCTCCAAAGTTTAC  1416

Oy      401  ProProValProAlaGluUysArgUysPro1IeArgValLeuSer1eUysPheAspGly1Ie  420
Db      1417  CCACCTGTCCAGCTGAAGAGAGAGAGCCCATCGGGTGTGTCTCTTTGATGGAATC  1476

Oy      421  AlaThrGlyLeuLeuVal1LeuUysAspLeuGly1IeGlnVal1AspArgTyr1IeAlaSer  440
Db      1477  GCTTACAGGCTCTGTGTGTGTAAGACTTGGGCACTTCAAGTGGACCGTTACATTCCTCG  1536

Oy      441  GluValCysGluAspSer1IeThrValGlyMetValArgHISGlnGlyUys1IeMetTyr  460
Db      1537  GAGGTGTGTAGGACTCATCAGCGTGGCATGTGTGGCCAGCAGGGAAATCATGTAC  1596

Oy      461  ValGlyAspValArgSerValThrGlnUysHIS1IeGlnGluTTPGlyProPheAspLeu  480
Db      1597  GTCGGGAGCGTCCAGCGTCAACAGAGCATATCCAGAGGTGGGGCCATTTGATTTG  1656

Oy      481  Val1IeGlyGlySerProCysAsnAspLeuSer1IeValAsnProAlaArgUysGlyLeu  500
Db      1657  GTGATTTGGGGCAATCCCTGCAATGACTCTTCCATCTGTCAACCTGTCCGAAGGCTTC  1716

Oy      501  TyrGlnGlyThrGlyArgLeuPheGluPheTyrArgLeuLeuHISAspAlaArgPro  520
Db      1717  TAGAGGGCACTGGCCGCGCTCTTGTGATTTACCGCCCTCGATGATGAGCCGCGCC  1776

Oy      521  LysGluGlyAspAspArgProPhePheTTPLeuPheGluAsnVal1ValAlaMetGlyVal  540
Db      1777  AAGAGGAGATGATCGCCCTTCTTGTGCTTTGAAGATGTGGTGGCCATGGGCGTT  1836

Oy      541  SerAspUysArgAsp1IeSerArgPheLeuGluUysSerAsnProValMet1IeAspAlaUys  560
Db      1837  AGTACAAAGAGGACATCTGCGCATTTCTCGAGTCCAAACCTGTGATGATGATGCCAA  1896

Oy      561  GluValSerAlaAlaHISArgAlaArgTyrPheTTPGlyAsnLeuProGlyMetAsnArg  580
Db      1897  GAATGTCAAGTGCACACAGAGGCCCGCTACTTGGGGTAACTTCCCGGTATGAACAG  1956

Oy      581  ProLeuAlaSerThrValaAsnAspUysLeuGluLeuGlnGluCysLeuGlnHISGlyArg  600
Db      1957  CCGTTTGATCTCAGTGAATGATGATGAGTGGAGGTGAGGTCTGTGGAGCATGGGCGG  2016

Oy      601  1IeAlaUysPheSerUysValArgThr1IeThrThrArgSerAsnSer1IeUysGlnGly  620
Db      2017  ATACCAAGTTTCAAGAAAGTGAAGCACTTACTHCGAGGTAAACTCATTAAGAGAGGC  2076

Oy      621  LysAspGlnHISpeProValPheMetAsnGluUysGluAsp1IeLeuTTPCysThrGlu  640
Db      2077  AAAGACCAAGCTTTCTGCTTCAATGATGAGAAAGGAGCACTTATGTGACACTGAA  2136

Oy      641  MetGluArgValPheGlyPheProValHISThrThrAspVal1SerAsnMetSerArgLeu  660
Db      2137  ATGAAAGGATTTGTTCCTCCAGTCACTATCTAGCTTCACATATGAGCGCGCTTG  2196

```

QY 661 AlaArglnuArgleuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
 Db 2197 GCGAGGCGAGACTGCTGGGCGGCTCAGGCGGTGCGACATCCGACCTCTTGCT 2256  
 QY 681 ProLeuYsgLurPheAlaCysVal 689  
 Db 2257 CCGCTGAAGAGATATTTGGCTGTGTG 2283  
 RESULT 2  
 US-10-144-577-2  
 ; Sequence 2, Application US/10144577  
 ; Publication No. US20030083292A1  
 GENERAL INFORMATION:  
 ; APPLICANT: Macleod, Alan Robert  
 ; TITLE OF INVENTION: Inhibitors of DNA Methyltransferase Isoforms  
 ; FILE REFERENCE: MET-005  
 ; CURRENT APPLICATION NUMBER: US/10/144,577  
 ; CURRENT FILING DATE: 2002-05-13  
 ; PRIOR APPLICATION NUMBER: US 60/290,202  
 ; PRIOR FILING DATE: 2001-05-14  
 ; PRIOR APPLICATION NUMBER: US 60/290,212  
 ; PRIOR FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 3005  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-144-577-2  
 Alignment Scores:  
 Pred. No.: 0 Length: 3005  
 Score: 3763.00 Matches: 689  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 Gaps: 0  
 US-10-623-813-86 (1-689) x US-10-144-577-2 (1-3005)  
 QY 1 MetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnuValGluGluAlaSer 20  
 Db 899 ATAAATGCTGTGAAGAAACCAAGGGCCCGGAGCTCTCAAGAGTGGAGAGGCCAGC 958  
 QY 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40  
 Db 959 CCTCTGCTGTGAGAGCCCACTGACCCCGCATCCCTGCTGCTACACGCTGAG 1018  
 QY 41 ProValGlySerAspAlaGlyAspAlaThrValAlaGlyAspAspGluProGlu 60  
 Db 1019 CCGGTGGGTCCATGCTGGGGACMAAGATGCCCAAGCGAGCGATGACAGCCAGAG 1078  
 QY 61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPhe 80  
 Db 1079 TACGAGGACGGCCGGGCTTTGGCATTTGGGAGCTGTTGGGGAAACTGCGGCTTC 1138  
 QY 81 SerTrpTrpProGlyArgGlyLeuAlaSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
 Db 1139 TCTGTGGTGGCCAGGCGCATTTGTCTTGGTGTGATGACGGGCGCGAGCCAGCAGCTGAA 1198  
 QY 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys 120  
 Db 1199 GGCACCCGCTGGGTCAATGAGTTCGGAGCGCAAAATTTCTAGTGTGTGTGTGAGAG 1258  
 QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140  
 Db 1259 CTATGCGCTGAGCTGCTTTTGGAGTGCCTTCCACAGCGCCAGTCAACAAGCACACCC 1318  
 QY 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 Db 1319 ATGTACCGCAAAAGCCATCTACAGAGTCTCTGACAGGTGCGACAGCCCGCGGGAAGCTG 1378

QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaValAlaGluValGlnAsn 180  
 Db 1379 TTCCCGGTGTGCCACGACGAGCGATGAGAGTGAACACTGCCAAGCCCGTGGAGGTGCACAA 1438  
 QY 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
 Db 1439 AACCCATGATTAATGGGCGCTGGGGGCTTCAGCTTCTGCGCTTCAAGGGGCTGGAG 1498  
 QY 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluPro 220  
 Db 1499 CCACCAAGAAAGAAAGAAATCCCTCAAGAAAGATGACAGGACATGTGGTGGAACTT 1558  
 QY 221 GluAlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
 Db 1559 GAGGAGCTGCTTACCGCACCACTCCACAGCCAAAAGCCCGGGAAGACACAGCGGAG 1618  
 QY 241 LysProLysValLysGluIleLeuAspGluArgThrArgGluArgLeuValTyrGluVal 260  
 Db 1619 AAGCCCAAGGTCAAGGAGATTATGTAGTGAAGCGCACAAAGAGCGGCTGTGTACGAGGTG 1678  
 QY 261 ArgGluLysCysValArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 Db 1679 GCGCAGAAAGTCCGGAACATTTGAGACATCTGCATCTCTGTGGAGCTTCATATGTACC 1738  
 QY 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
 Db 1739 CTGGAACACCCCTCTTCTGTGGAGGAAATGTGCCAAAACGCAAGAACTCTTCTTGAG 1798  
 QY 301 CysAlaTyrGlnTyrAspAspArgGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
 Db 1799 TGTGCTACCAAGTACGACGACGAGCGCTACAGCTCTACTGACCACTCTGCTGGGGGC 1858  
 QY 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
 Db 1859 CGTGAAGTGTCTATGTGGGAAACAACACTGCTGAGGCTTTTGGGTGGAGTGTGTG 1918  
 QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyrAsnCys 360  
 Db 1919 GACCTCTGTGGTGGGCGGGGCTGCCAGGAGCCATTAAAGAAAGCCCTGGAAGCTGC 1978  
 QY 361 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgAlaAspTrpProSer 380  
 Db 1979 TACATGTGCGGACAAAGGTACTTACGGGCTGCTCGGGGCGAGAGGACTGGCCCTCC 2038  
 QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProPolysValTyr 400  
 Db 2039 CGGCTCCAGATGTTCTTCCCTAATAACACAGCCAGGAATTTCACCTCCAAAGTTTAC 2098  
 QY 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
 Db 2099 CACCTGTCCCACTGAGAAAGAAAGCCCATCGGGTGTCTCTCTTGAATGGAATC 2158  
 QY 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
 Db 2159 GCTACAGGGCTCTGTGTGTAAGGACTTGGGCATTCAGGTGACCGCTCATTTGCCCTCG 2218  
 QY 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
 Db 2219 GAGGTGTGTAAGACTCCACACAGTGGGATGTGGGACCAAGGGGAATCATATAC 2278  
 QY 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeu 480  
 Db 2279 GTGCGGAGCTCCGACGCTGACACAGAAAGCATATCCAGGAGTGGGGCCATTGCAATCTG 2338  
 QY 481 ValIleGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db 2339 GTGATTTGGGGGAGTCCCTGCAATGACCTTCCATGTCAACCTGTCACAAAGGGCTC 2398  
 QY 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
 Db 2399 TACGAGGACCTGGCGGCTCTTTTGAATTTACCGCTCTCGCATGATGGCGGCGCC 2458  
 QY 521 LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyVal 540

Db 2459 AAGGAGGAGATGATCCGCTTTCTTCTGCTCTTTAGATGATGAGGCGCATGGGCTT 2518  
Qy 541 SerAspLysArgAspLLeSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
Db 2519 AGTGACAAGAGGAGCATCTGGCATTTCTCGAGTCCAACCTGTGATGATGATGCCAAA 2578  
Qy 561 GluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg 580  
Db 2579 GAAGTGTACGTGCACACAGGCGCGCTACTCTGGGGTAACTTCCCGGTATGAACAGG 2638  
Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArg 600  
Db 2639 CCGTTGGCATCTCACTGTGAATGATAAAGCTGAGCTGCAGAGGTCTTCGAGCATGGCAGG 2698  
Qy 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
Db 2699 ATAGCCAAAGTTCAACAAAGTAGACACCATTTACTACGAGGTCAAACTCCATAAACAGGGC 2758  
Qy 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGlu 640  
Db 2759 AAGAGCCAGCATTTTCTGTCTTCTATGATAGAAAGAGGACATCTATGGTGACCTGAA 2818  
Qy 641 MetGluArgValPheGlyPheProValHisTrpThrAspValSerAsnMetSerArgLeu 660  
Db 2819 ATGGAAAGGATATTGGTTTCCAGTCCACTATTACTGACGTCTCCAAACATGAGCCGCTTG 2878  
Qy 661 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
Db 2879 GCGAGGACAGACACTGCTGGCGCGGTCACTGAGCGGTGACATCCGCGCACTCTCGCT 2938  
Qy 681 ProLeuLysGluTyrPheAlaCysVal 689  
Db 2939 CCGCTGAAGAGTATTGTGCTGTGTG 2965

## RESULT 3

US-10-172-118-132  
/ Sequence 132, Application US/10172118  
/ Publication No. US20030224374A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Dai, Hongyue  
/ APPLICANT: He, Yundong  
/ APPLICANT: Linsley, Peter  
/ APPLICANT: Mao, Mao  
/ APPLICANT: Roberts, Chris  
/ APPLICANT: Van 't Veer, Laura  
/ APPLICANT: Van de Vijver, Marc  
/ APPLICANT: Bernards, Rene  
/ TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
/ FILE REFERENCE: 9301-175-999  
/ CURRENT APPLICATION NUMBER: US/10/172,118  
/ CURRENT FILING DATE: 2002-06-14  
/ PRIOR APPLICATION NUMBER: 60/380,770  
/ PRIOR FILING DATE: 2002-05-14  
/ NUMBER OF SEQ ID NOS: 2699  
/ SEQ ID NO 132  
/ LENGTH: 3005  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ PUBLICATION INFORMATION:  
/ DATABASE ACCESSION NUMBER: AF067972  
/ DATABASE ENTRY DATE: 2001-06-18  
/ US-10-172-118-132

## Alignment Scores:

Pred. No.: 0 Length: 3005  
Score: 3763.00 Matches: 689  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 7 Gaps: 0

US-10-623-813-86 (1-689) X US-10-172-118-132 (1-3005)

Qy 1 MetAsnAlaValGluGluLeuGlnGlnGlyProGlyGlySerGlnLysValGluGluAlaSer 20  
Db 899 ATAAATGCTGTGAAAGAAACCAAGGAGCCCGGGAGTCTCAAGAAAGTGGAGAGGCCAGC 958  
Qy 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40  
Db 959 CTTCCGTGCTGGACAGAGCCCACTGAACCCCGCATCCCGCATCTGGGTCAACAGCCCTGAG 1018  
Qy 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
Db 1019 CCGGTGGGGTCCCATGCTGGGGAACAAGAACCAACCAACAGCCGATGACGAGCCAGAG 1078  
Qy 61 TyrGluAspGlyValArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPhe 80  
Db 1079 TAGAGAGACCGCGGGGCTTTGGCATTTGGGAGCTGTGTGGGGGAACCTCCGGGCTTC 1138  
Qy 81 SerTrpTrpProGlyValArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 100  
Db 1139 TCTGGTGGCCAGGCGGCATTTGTCTTGGTGGATGACGGGCGGAGCCGAGAGCTGAA 1198  
Qy 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGlnLys 120  
Db 1199 GGCACCGCTGGGTCAATGTGTGAGAGAGCGCAAAATCTCAGTGTGTGTGAGAGAG 1258  
Qy 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140  
Db 1259 CTGATGCCGCTGAGCTGCTTTTTCGATGCTTCCACCAAGCCACGTCACAAAGCACACCC 1318  
Qy 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
Db 1319 ATGATCCGCAAAACCAATCTACGAGGTCTCTCAGGTGGCCAGAGCCGCGGGGAAACTG 1378  
Qy 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValAsn 180  
Db 1379 TTCCCGGTGTGCACACACAGCATGAGAGTGAACCTCCAAAGCCGAGGAGTGCAGAAC 1438  
Qy 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
Db 1439 AAGCCCATGATTGAATGGGCTTGGGGGCTTCCAGCTCTTGGCCCTTAAGGGCTTGAG 1498  
Qy 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGlnPro 220  
Db 1499 CCAACGAAAGAGAAAGAAATCCCTCAAGAAAGTGAACAGACATGTGGTGAACCT 1558  
Qy 221 GluAlaAlaAlaTyrAlaProProProProAlaLysSerProArgLysSerThrAlaGly 240  
Db 1559 GAGGAGCTCTTACGACACCACTCCACCAAGCCAAAGCCCGGAAGAGACACGCGAG 1618  
Qy 241 LysProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluVal 260  
Db 1619 AAGCCCAAGGTCAAGAGATTAATTGATGACGCAAGAGAGCGGTGTGACGAGGTG 1678  
Qy 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
Db 1679 GCGGAGAGTCCGGAACATTGAGGACATCTGATCTCTGTGGAGACCTCAATGTTACC 1728  
Qy 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
Db 1739 CTGGAAACCCCTCTTCTGTGGAGGATGTGCCAAACCTGCAGAACTCTTCTTGAG 1798  
Qy 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
Db 1799 TGTGCGTACAGTACAGACAGACGAGGTACCAATCTCACTGACCACTCTGTGGGGGC 1858  
Qy 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
Db 1859 CGTAGAGTGTCTATGTGCGGAAACCAACACTGTGCGAGGTCTTTTCCGTGAGTGTGTG 1918  
Qy 341 AspLeuLeuValGlyProGlyValaGlnAlaAlaIleLysGluAspProTrpAsnCys 360  
Db 1919 GACCTTGTGTGGGCGGGGGCTGCCAGGACCATTAAGAGAACCCCTTGAACCTGC 1978

```

Qy 361 TyMetCyseGlyHslySGlyThrTyrGlyLeuLeuArgArgGlyLysArgTyrProSer 380
Db 1979 TACTGTGCGGGGCAAGGGGCTTACCGGCTCTCGCGCGGCGAGAGACTGGCCCTCC 2038
Qy 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400
Db 2039 CGGCTCCAGATGTTCTTCGTAATATACCAACAGACAGAAATTGACCTCCAAAGGTTTAC 2098
Qy 401 ProProValProAlaGlyLysArgLysProLysArgValLeuSerLeuPheAspGlyTle 420
Db 2099 CCACCTGTCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 2158
Qy 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyTleGlnValAspArgTyrTleAlaSer 440
Db 2159 GCTACAGAGGCTCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2218
Qy 441 GluValCysGlnAspSerLethrValGlyMetValArgHisGlnGlyLysLysLysLysLys 460
Db 2219 GAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 2278
Qy 461 ValGlyAspValArgSerValThrGlnLysHisLleGlnGlyTyrProPheAspLeu 480
Db 2279 GTCGGGAGAGCTCCGACAGCTCACAGAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGATC 2338
Qy 481 ValLleGlyLysSerProCysAsnAspLeuSerLleValAsnProAlaArgLysGlyLeu 500
Db 2339 GTATATGGGGGAGTCCCTGAAATGACCTTCCATTCGCAACCTTGCTCGAGAGAGAGCTTC 2398
Qy 501 TyrGlnGlyThrGlyArgLeuPhePheGlnPheTyrArgLeuLeuHisAspAlaArgPro 520
Db 2399 TAGAGAGGAGCTGGCGCGCTCTTCTTGAGTTTACCGGCTCCCTCGCATGATGGCGGCCC 2458
Qy 521 LysGlnGlyAspAspArgProPhePheTyrLeuPheGlnLysValLleAlaMetGlyVal 540
Db 2459 AAGGAGGAGATGATCGCCCTTCTTGCTGCTTGTGAGATGTGGTGGCCATGGGGGCTT 2518
Qy 541 SerAspLysArgAspLysArgPheLeuGlnSerAsnProValMetLleAspAlaLys 560
Db 2519 AGTGACAGAGGAGCATCTCGCGATTTCTCGAGTCAACCTGTGATGATGTATGCCAA 2578
Qy 561 GluValSerAlaAlaHisArgAlaArgTyrPheTyrPheGlyAsnLeuProGlyMetAsnArg 580
Db 2579 GAAATGCTCACTGCACACAGGGCCGCTACTTCTGGGTTAACCTTCCCGGATGAACAGG 2638
Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGlnGlyLysLeuGlnHisGlyArg 600
Db 2639 CCGTTGGCATCTCACTGAAATGATTAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2698
Qy 601 LLeAlaLysPheSerLysValArgThrLleThrThrArgSerAsnSerLleLysGlnGly 620
Db 2699 ATAGCCAAAGTTCAAGCAAGTGAAGCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2758
Qy 621 LysAspGlnHisPheProValPheMetAsnGlnLysGlnAspLysLeuTyrPyrGln 640
Db 2759 AAAGACACAGACTTTCTGTCTTTCATGAATGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 2818
Qy 641 MetGlnArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660
Db 2819 ATGGAAGAGGATTTGGTTTCCAGTCCACTATACGTCGCTCCAAACATGAGAGCGGCTTG 2878
Qy 661 AlaArgGlnArgLeuGlnArgSerTyrSerValProValLleArgHisLeuPheAla 680
Db 2879 GCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2938
Qy 681 ProLeuLysGlnTyrPheAlaCysVal 689
Db 2939 CCGCTGAAGAGATTTTGCCTGTGTG 2965

```

RESULT 4  
 US-10-342-887-132  
 ; Sequence 132, Application US/10342887  
 ; Publication No. US20040058340A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Robert, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 132
; LENGTH: 3005
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-132
Alignment Scores:
Pred. No.: 0 Length: 3005
Score: 3763.00 Matches: 689
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
Gaps: 0
US-10-623-813-86 (1-689) x US-10-342-887-132 (1-3005)
Qy 1 MetAsnAlaValGlnGluAsnGlnGlyProGlyGlnSerGlnLysValGlnGlnAlaSer 20
Db 899 ATGATGCTGTGAGAAAGAAACAGAGGGGCGGGGAGTCTCAGAAAGTGAGAGAGCCAGC 958
Qy 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGln 40
Db 959 CCTCTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
Qy 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGlnProGln 60
Db 1019 CCGTGGGCTCCATGCTGTGGGAGCAAGATGCCAACAGAGCCATGACAGCCAGAG 1078
Qy 61 TyrGlnAspGlyArgGlyPheGlyLleGlyGlnLeuValTyrGlyLysLeuArgGlyPhe 80
Db 1079 TAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138
Qy 81 SerTyrTyrProGlyArgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
Db 1139 TCTGTGTGGCCAGAGCGGCAATGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1198
Qy 101 GlyThrArgTyrValMetTyrPheGlyAspGlyLysPheSerValValCysValGlnLys 120
Db 1199 GGGAGCGGCTGGGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1258
Qy 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140
Db 1259 CTGATGCGGCTGAGAGCTGTTTTCAGATGAGCTTCCACAGGCGCAGATCAACAGCAGCCC 1318
Qy 141 MetTyrArgLysAlaLysTyrGlnValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160
Db 1319 ATGTACCGCAAGACCACTCAAGAGTCTCAGAGTCCAGAGTCCAGAGCCCGGGGAGAACTG 1378
Qy 161 PheProValCysHisAspSerAspGlnSerAspThrAlaLysAlaValGlnValAsn 180
Db 1379 TTCCCGGTGTGCCACAGAGAGATGAGAGTGAAGTCACTGCCAAGCGCGTGAAGTCAAGAC 1438
Qy 181 LysProMetLleGlnTyrPalLeuGlnGlyPheGlnProSerGlyProLysGlyLeuGln 200

```

Db 1439 AAGCCCATGATTGATGGGCCCTTGGGGGCTTCCAGCCTTCTGGCCCTTAAGGCCCTGGAG 1498  
 Qy 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTyrValGluPro 220  
 Db 1499 CCACCAGAAAGAGAGAAATATCTCTCAAAAGAGTGTACACGACATGTGGGTGAGACCT 1558  
 Qy 221 GGUUUAUAAATATYrValAProProProProAlaLysLysProAlaLysSerThrLaglu 240  
 Db 1559 GAGGAGCTGCTTACGACACCTCCACGACGAAAAAGCCCCGGAGAGGACACAGCGGAG 1618  
 Qy 241 LysProLysValLysGluLeileLysArgLysArgLysArgLysValTyrGluVal 260  
 Db 1619 AAGCCCAAGGTCAAGAGATTATGTAGCGCCACAGAGAGCGGCTGTGTACAGAGTG 1678  
 Qy 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 Db 1679 CGGAGAAAGTCCGGAAACATTGAGACATCTGCATCTCTGTGGAGGACCTCATTTTACC 1738  
 Qy 281 LeuGlnHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
 Db 1739 CTGGAACACCCCTCTTCTGTGGAGAAATGTGCCAAACCTGCMAACACTCTTTCTGGAG 1798  
 Qy 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGly 320  
 Db 1799 TGTGCGTACCAAGTACGACGACGAGCGGCTACCAAGTCTTACCATCATCTGTGGGGC 1858  
 Qy 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
 Db 1859 CGTGAGGTCTCATGTGCGGAAACAAACACTGCTCAGAGTGTGTTGGTGGAGTGTG 1918  
 Qy 341 AspLeuLeuValGlyProGlyValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 360  
 Db 1919 GACCTCTGTGGGGCCGGGGGCTCCAGCGACGCAATTAAAGAAACCCCTCGAACTGC 1978  
 Qy 361 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTyrProSer 380  
 Db 1979 TACATGTGCGGGCAACAGGTACTTACGGGCTGTGCGGGCGGAGAGACTGGCCCTCC 2038  
 Qy 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysAlaTyr 400  
 Db 2039 CGGCTCCAGATGTTCTTCTGCTAAATCAACACGACCGGAATTTGACCTCCAAAGGTTTAC 2098  
 Qy 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyLe 420  
 Db 2099 CCACCTGCTCCAGCTGAGAAAGAGAAAGCCCATCGGGTGTCTCTCTTTGATGATGCATC 2158  
 Qy 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAspSer 440  
 Db 2159 GCTACAGGGGCTCTGTGCTCTGAAAGACTTGGGCATTCAAGTGGACCGCTACATTGCTCG 2218  
 Qy 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
 Db 2219 GAGGTGTGTGAGACTCATACGCTGGGCATGTGGCGGACACAGGGAGAAATCATGTGAC 2278  
 Qy 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeu 480  
 Db 2279 GTGGGGAAGCTCCGAGGCTCAACAGAAACATATCCAGAGTGGGCCCATTCATCTG 2338  
 Qy 481 ValIleGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db 2339 GTGATTTGGGGGAGTCCCTGCAATGACTTTCATCGTCAACCTGTGCGCAAGGGCTTC 2398  
 Qy 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
 Db 2399 TACAGGGGCACTGGCGGCTCTTCTTGAATTCATCCGCTTCGCAAGATGGCGGGCTTC 2458  
 Qy 521 LysGluGlyAspAspArgProPhePheThrLeuPheGluAsnValValAlaMetGlyVal 540  
 Db 2459 AAGAGGAGAGTATGATCGCCCTCTTCTGTGGCTCTTGAAGATGTGTGGCCATGGGCTTC 2518  
 Qy 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
 Db 2519 AGTGACAAAGAGGACATCTCGGATTTCTCGAGTCCAACCTGTGATGATGATGATGCCAAA 2578

Qy 561 GluValSerAlaAlaAlaHisArgAlaArgTyrPheThrGlyAsnLeuProGlyMetAspArg 580  
 Db 2579 GAAGTGTCACTGTGACACAGGGCCCGCTACTTCTGGGGTAACTTCCGGATATGAACAGG 2638  
 Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGlnHisGlyArg 600  
 Db 2639 CCGTTGGCATCTGTGAATGATGAATTAAGTGGAGTGTGACGAGTGTCTGGAGCATGGCAGG 2698  
 Qy 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
 Db 2699 ATACCCAAAGTTCACACAAAGTGAAGCATTACTACGAGGTCAAACTCCATTAACAGGGC 2758  
 Qy 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGlu 640  
 Db 2759 AAGACCAAGCTTTTCTGTCTTATGAAGAAGAAAGACATCTTATGGATGACTGAA 2818  
 Qy 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
 Db 2819 ATGGAAGGCGTATTTGGTTTCCAGTCCAGTCACTATGACGTCTCCAAACATGAGCGGCTTG 2878  
 Qy 661 AlaArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAla 680  
 Db 2879 GCGAGGAGAGACTGTGCGCGGCTCATGAGCGTGCACATCATCCGACCTTCTGCT 2938  
 Qy 681 ProLeuLysGluTyrPheAlaCysVal 689  
 Db 2939 CCGCTGAAGAGATTTTGTGCTGTGTG 2965  
 RESULT 5  
 US-10-330-773-279  
 ; Sequence 279, Application US//10330773  
 ; Publication No. US2006040262A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: David W. Morris  
 ; APPLICANT: Marc Malandro  
 ; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
 ; FILE REFERENCE: 529452001300  
 ; CURRENT APPLICATION NUMBER: US//10/330,773  
 ; NUMBER OF SEQ ID NOS: 981  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 279  
 ; LENGTH: 3009  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-330-773-279  
 Alignment Scores:  
 Pred. No.: 0 Length: 3009  
 Score: 3763.00 Matches: 689  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 11 Gaps: 0  
 US-10-623-813-86 (1-689) x US-10-330-773-279 (1-3009)  
 Qy 1 MetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAlaSer 20  
 Db 903 ATGAATCTGTGGAAAGAAACAGAGGGGCCGGGAGTCTCGAAGGTGGAGAGGCCAGC 962  
 Qy 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40  
 Db 963 CCTCTCTGTGGGACAGCCCACTGACCCGCAATCCCACTGTGGCTTACCAAGCTGTAG 1022  
 Qy 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
 Db 1023 CCGTGGGGTCCGATGTCTGGGGAACAAGATGCCCAAGAGCGATGACGAGCCAGAG 1082  
 Qy 61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPhe 80  
 Db 1083 TACAGAGACGGCCGGGCTTTGGCATTTGGGGAAGCTGTGTGTGGGGAACCTGCGGGCTTC 1142

QY 81 SerTrpTrpProGluValArgIleValSerTrpTrpMetThrGlyArgSerArgAlaIleGlu 100  
 Db 1143 TCCTGGTGGCCAGGCGCATGTGTGTTGGTGGAGAGCGGCGGAGCGGACCTGAA 1202  
 QY 101 G1YThrArgTrpValMetTrpPheGlyValSerPheSerValValCysValGluLys 120  
 Db 1203 GGCAACCGCTGGGTATGTGGTTCGGAGACGGCAAAATCTCAGTGGTGTGTGGAGAG 1262  
 QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrThrAsnLysGlnPro 140  
 Db 1263 CTGATGCGCGCTGAGCTCTTTTTCAGTCCGTTCCACGAGCCACGTAAACACAGCAGCC 1322  
 QY 141 MetTrpArgLysAlaIleTrpGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 Db 1323 ATGTACCGCAAGCCATCTACAGAGTCTCTGACAGTGGCGACGCGCGGGGAAAGCTG 1382  
 QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsn 180  
 Db 1383 TTCCCGGTGTCCACGACGAGCATGAGATGACACTGCCAAGCCGTGGAGGTGCAGAAC 1442  
 QY 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
 Db 1443 AAGCCCATGATTGAATGGGCGCTGGGGGCTTCCAGCCTTCTGGCCCTAAGGGCGTGGAG 1502  
 QY 201 ProProGluGluGluLysAsnProGlyTrpGlyValAlaThrTrpAspMetTrpValGluPro 220  
 Db 1503 CCACCAAGAAAGAAAGAAATCCCTCAAAAGAAAGTACAGCGCATGTGGTGGAACTT 1552  
 QY 221 GluAlaAlaAlaValAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
 Db 1563 GAGGCACTGTGCTTACGACCACTCCACACCAAAAGCCCCGGAAAGACAGCCGAG 1622  
 QY 241 LysProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTrpGluVal 260  
 Db 1623 AAGCCCAAGGTCAAGGAGATTGATGACGCGCACAGAGAGCGGCTGTGTACGAGGTG 1682  
 QY 261 ArgGlnLysCysAspGlnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 Db 1683 CGGAGAAAGTGCCTGGAACTTGAAGACATCTGCACTCTGTGGAGGCTTCAAGTTTACC 1742  
 QY 281 LeuGlnHisProLeuPheValGlyLysCysGlnAsnCysLysAsnCysPheLeuGlu 300  
 Db 1743 CTGGAACACCCCTCTTCTGTGGAGGAAATGTGCCAAACTGCAGAACTGCTTCTGGAG 1802  
 QY 301 CysAlaArgTrpGlnTrpAspAspAspGlyTrpGlnSerTrpCysThrIleCysCysGlyGly 320  
 Db 1803 TGTGCGTACCAAGTACGACGAGCAGCGGCTACCAAGTCTACTGCACATCTGCTGTGGGCG 1862  
 QY 321 ArgGluValLeuMetCysGlyAsnAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
 Db 1863 CGTGAGGTGCTCATGTGGGAAACAAACAACGTGCAAGGTCTTTTGGTGGAGTGTGTG 1922  
 QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTrpAsnCys 360  
 Db 1923 GACTCTTGGTGGGGCGGGGGCTGCCAGGACCCATTAGAAAGACCCCTTGGAACTGC 1982  
 QY 361 TyrMetCysGlyHisLysGlyThrThrGlyLeuLeuArgArgGluAspTrpProSer 380  
 Db 1983 TACATGTGGGGGCAACAAGGTACTACGGGCTGCTGGCGGGCGGAGAGACTGGCCCTCC 2042  
 QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400  
 Db 2043 CGGCTCCAGAGTCTTCTGCTAAATTAACAACAAGAAATTGACCTTCCAAAGGTTTAC 2102  
 QY 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
 Db 2103 CCAACCTGTCCAGTGTGAAGAGAGAACCCCATCCGGGTGCTGTCTCTTTGATGAGATC 2162  
 QY 421 AlaThrGlyLeuLeuValLysLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
 Db 2163 GCTACAGAGGCTCCTGTGTGTGAAGAGACTTGGGCACTTCCAGTGGAGCCGCTACATGCTCG 2222

QY 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTrp 460  
 Db 2223 GAGGTGTGTGAGACTTCATACCGTGGCATGTGGTCCGGCCACGAGGAAATCATGTAC 2282  
 QY 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeu 480  
 Db 2283 GTCCGGGACGTCCCGACAGGTCAACAAGACATATCCAGAGTGGGGCCATTGCATCTG 2342  
 QY 481 ValIleGlyLysProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db 2343 GTGATTTGGGGCACTCCCTGCAATGACCTTCCATGCTCACTGCAACCTGTCCGAAAGGCTTC 2402  
 QY 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTrpArgLeuLeuHisAspAlaArgPro 520  
 Db 2403 TACAGGGCACTGGCGGCTCTTCTTGAGTTCTACCGGCTTCTGCAATGACGCGGCC 2462  
 QY 521 LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValAlaMetGlyVal 540  
 Db 2463 AAGGAGGAGATGATCCGCCCTTCTTGGCTTTTGAAATGTGTGGCCATGGGCGTT 2522  
 QY 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
 Db 2523 AGTACAAAGAGGACATCTGCGCATTTCTGAGTCCAAACCTGTGATGATTGATGCCAA 2582  
 QY 561 GluValSerAlaAlaHisArgAlaArgTrpPheTrpGlyAsnLeuProGlyMetAsnArg 580  
 Db 2583 GAGGTGTACGTGCACACAGGCGGCTTCTTGGGGTAACTTCCCGGTAATGAAACGG 2642  
 QY 581 ProLeuLaserThrValAsnAspLysLeuGluLeuGlnGlyCysLeuGlnHisGlyArg 600  
 Db 2643 CCGTTGCATTCACCTGGAATGATGACTGAGGTGCAGAGGTGTCTGAGACATGGCAGG 2702  
 QY 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
 Db 2703 ATACCCAAAGTTCAGCAAAAGGAGACCATTAACAAGGTCAACCTCCATAAACAGAGGC 2762  
 QY 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGlu 640  
 Db 2763 AAGACACGACATTTTCTGTCTTCTCATGAAATGAGAAAGACATCTTATGGTGCACGAA 2822  
 QY 641 MetGluArgValPheGlyPheProValHisTrpThrAspValSerAsnMetSerArgLeu 660  
 Db 2823 ATGGAAGGGATTTGGTTTCCAGTCCACTATGACAGCTTCACATGACGCGCTTG 2882  
 QY 661 AlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAla 680  
 Db 2883 GCGAGGCAAGACTGCTGGGCGGCTCATGAGCGGTGCCAGTCAATCCGCACCTTCTGCT 2942  
 QY 681 ProLeuLysGluTrpPheAlaCysVal 689  
 Db 2943 CCGCTGAAGAGTATTTTGGCTGTGTG 2969

RESULT 6  
 US-09-720-086-3  
 : Sequence 3, Application US/09720086  
 : Publication No. US20060084053A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Li, En  
 : APPLICANT: Okano, Masaki  
 : APPLICANT: Xie, Shaoqing  
 : TITLE OF INVENTION: De Novo DNA Cytosine Methyltransferase Genes,  
 : TITLE OF INVENTION: Polypeptides & Uses Thereof  
 : FILE REFERENCE: 0609,4560002  
 : CURRENT APPLICATION NUMBER: US/09/720,086  
 : CURRENT FILING DATE: 2000-12-20  
 : PRIOR APPLICATION NUMBER: PCT/US99/14373  
 : PRIOR FILING DATE: 1999-06-25  
 : PRIOR APPLICATION NUMBER: 60/090,906  
 : PRIOR FILING DATE: 1998-06-25  
 : PRIOR APPLICATION NUMBER: 60/093,993  
 : PRIOR FILING DATE: 1998-07-24  
 : NUMBER OF SEQ ID NOS: 82  
 : SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 3
; LENGTH: 4293
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-720-086-3

Alignment Scores:
Pred. No.: 0
Score: 3763.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 100.0%
DB: 5

Length: 4293
Matches: 689
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-623-813-86 (1-689) x US-09-720-086-3 (1-4293)

Qy      1 MetAenAlaValIGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAsnSer 20
Db      899 ATGAATGCTGTGGAGAAACCAAGGGGCGCGGGAGTCTCAAGAGTGGAGAGGCCAGC 958

Qy      21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrProGlu 40
Db      959 CCTCTGCTGTGCACAGCCCACTGACCCCGCATCCCGCACTGTGGTCAACCGCTGAG 1018

Qy      41 ProValIGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60
Db      1019 CCGGTGGGGTCCGATGTGGGGACAAGATGCCCAAGCGAGCGATGACGAGCCAGAG 1078

Qy      61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValITripGlyLysLeuArgGlyPhe 80
Db      1079 TACAGAGACGGCGGGGCTTTGGCATTTGGGAGCTGTGTGGGGAAACTGGGGGCTTC 1138

Qy      81 SerITripProGlyArgGlyLeuIleSerITripMetThrGlyArgSerArgAlaIleGlu 100
Db      1139 TCCTGTGGGCCCGCGCATTTGTCTTGTGTGATGACGGCGCGGACGACACTGAA 1198

Qy      101 GlyIThrArgITripValMetITripPheGlyLysPheSerValValCysValGluLys 120
Db      1199 GGCACCCCTGGGTATGTGGTTCGAGACGGCAAAATCTCAGTGTGTGTGAGAAAG 1258

Qy      121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140
Db      1259 CTGATGGCGCTGAGCTCGTTTGGAGTGGGTTCCACAGGCCACGTAACAAGAGGCC 1318

Qy      141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160
Db      1319 ATGTACCCCAAGCATCTACAGAGTCTCTGCGAGTGGCCAGCAGCCGCGGGAAAGCTG 1378

Qy      161 PheProValCysHisAspSerAspGluSerAspITripAlaLysAlaValGluValAsn 180
Db      1379 TTCCGGGTGTCCAGACAGCATAGATGACACTGGCCAGGCCGTGAGAGTGCAGAAC 1438

Qy      181 LysProMetIleGluITripAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200
Db      1439 AAGGCCAAGATTGAATGGCCCTGGGGGCTTCCAGCCCTTGGCCCTTAAGGGCTTGAG 1498

Qy      201 ProProGluGluGluLysAsnProTyrLysGluValITyrIThrAspMetITripAlaGluPro 220
Db      1499 CCACCAAGAAAGAGAAATCCCTCAAGAAAGTGTACACGAGCATGTGGGTGAAACCT 1558

Qy      221 GluAlaAlaAlaITyrAlaProProProProAlaLysLysProAlaGlySerIThrAlaGlu 240
Db      1559 GAGGAGGCTGCTTACGACACCATCTCACCAAGCAAAAGCCCCGGGAAGAGCACAGCGAG 1618

Qy      241 LysProLysValLysGluIleIleAspGluArgIThrArgGluArgLeuValITyrGluVal 260
Db      1619 AAGCCAAAGGTCAAGGAGATTATGATGAGCGCAACAAGAGCGGCTGTGTACAGAGTG 1678

Qy      261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValIThr 280
Db      1679 CGGCAAGAGTCCCGACATTTGAGAGACATCTGCACTCTCTGTGGAGAGCCCTCAATTTTACC 1738

Qy      281 LeuGluHisITripLeuPheValGlyLysCysGlnAsnCysLysAsnCysPheLeuGlu 300

```

```

Db      1739 CTGAAACACCCCTTCTTGTGTGGAGATGTGCCAAACTCCAAAGAACTCTTTCTGGAG 1798

Qy      301 CysAlaTyrGlnITyrAspAspAspGlyTyrGlnSerTyrCysIleCysCysGlyGly 320
Db      1799 TGTGTGTACCAAGTACAGACAGACGAGCTTACCGAGTCTTACCACTCTGTGTGGGGCC 1858

Qy      321 ArgGluValIleLeuMetCysGlyAsnAsnAsnCysArgCysPheCysValGluCysVal 340
Db      1859 GGTAGAGGCTCCAGTGTGGGAAACAAACATGTGACAGTGTCTTTGTGGAGATGTGTG 1918

Qy      341 AspLeuLeuValIGlyProGlyAlaIleGlnAlaIleLysGluAspProITripAsnCys 360
Db      1919 GACCTCTTGTGGGGCGGGGGCTGCCAGGACACCATTAAGAAAGACCCCTGGAATCTGC 1978

Qy      361 TyrMetCysGlyHisLysGlyIThrTyrGlyLeuLeuArgArgGluAspITripProSer 380
Db      1979 TACATGTGCGGGCAACAAGGTAACCTACGGCTGTGGCGCGGCGAGAGACTGGCCCTTC 2038

Qy      381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProITripValITyr 400
Db      2039 CGGCTCCAGATGTTCTTGTCTTAATTAACACACAGCAAGAAATTTGACCTCCAAAGGTTTAC 2098

Qy      401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420
Db      2099 CCACCTGTCCAGCTGAGAAAGAGAACCCATCCGGGTGCTGTCTCTTGTGATGAAATC 2158

Qy      421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValaAspArgTyrIleAlaSer 440
Db      2159 GCTACAGGGCTCTGTGTGTGAAGACTTGGGACTTCAAGGTGAGACCGCTTACATTTGCC 2218

Qy      441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460
Db      2219 GAGGTGTGTGAGACTCATTCACGCTGAGCATGTGTGGGCCCAAGGGAAATCATGTAC 2278

Qy      461 ValGlyAspValaArgSerValIThrGlnLysHisIleGlnGluITripGlyITripPheAspLeu 480
Db      2279 CTCGGGAGACGTCCGACAGCTCACACAGACATTCACAGAGTGGGCCCATTTGACTGTG 2338

Qy      481 ValIleGlyLysITripProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500
Db      2339 GTGATTTGGGGCAAGTCTCTGCAATGACCTCTCCATCTCTCAACCTCTGCGCAAGGGCTTC 2398

Qy      501 TyrGlnGlyIThrGlyArgGluPhePheGluPheTyrArgLeuLeuHisAspAlaArgPro 520
Db      2399 TACAGGGCACTGGCCGGCTCTTTTGAAGTCTTACCGGCTCTCTGCAAGAGCCGGGCC 2458

Qy      521 LysGluGlyAspAspArgProPhePheITripLeuPheGluAsnValAlaIleMetGlyVal 540
Db      2459 AAGGAGGAGATGATCCGCCCTTCTTGTGGCTCTTTGAGAAATGTGTGGCCATGGGGCTT 2518

Qy      541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560
Db      2519 AGTGACAAAGAGGACATCTCGCATTTCTCGAGTCCAAACCTGTGATGATGATGAGCCAAA 2578

Qy      561 GluValSerAlaAlaHisArgAlaArgTyrPheITripLysLeuPProGlyLysMetAsnArg 580
Db      2579 GAAATGTCAAGTGCACACAGGGCCCGCTTACTTCTGGGGTAAACCTTCCCGGATGAACAGG 2638

Qy      581 ProLeuAlaSerIThrValaAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArg 600
Db      2639 CCGTTGGCATTCACACTGTGAATGATAGCTGAGCTGCAAGAGATGTCTCGAGACATGGCAGG 2698

Qy      601 IleAlaLysPheSerLysValArgIThrIleIThrIThrArgSerAsnSerIleLysGlnGly 620
Db      2699 ATAGCCAAAGTTACAGCAAGTGAAGACCATTACTACGAGTCCAAACCTCCATAAAGCAGGGC 2758

Qy      621 LysAspGlnHisPheProValaPheMetAsnGluLysGluAspIleLeuITripCysIThrGlu 640
Db      2759 AAGACCAAGCATTTCTGTCTTATGAATGAGAAAGGACATCTTATGGTGCACCTGAA 2818

Qy      641 MetGluIThrValPheGlyPheProValaHisITyrIThrAspValSerAsnMetSerArgLeu 660

```

Db 2819 ATGAAAGGATATTGGTTCCTCCAGTCCACTATACTAGTCTCCCAACATGAGCCGCTTG 2878  
Qy 661 AAlaArgGlnAlaLeuGlnLeuGlyAArgSerThrSerValProValIleArgHisLeuPheIa 680  
Db 2879 GCGGAGGAGACATGCTGGCGGCTCATGAGCGCTGCAGTCACTCCGACCTTCCTGCT 2938  
Qy 681 ProLeuYsgGluTyrrPheAlaCysVal 689  
Db 2939 CCGCTGAAGAGATATTTTGCCTGTCTG 2965  
RESULT 7  
; Sequence 3, Application US/10623813  
; Publication No. US20040234997A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, En  
; APPLICANT: Okano, Masaki  
; APPLICANT: Xie, Shaoping  
; APPLICANT: Chen, Taiding  
; TITLE OF INVENTION: De Novo DNA Cytosine Methyltransferase Genes, Polypeptides & Uses  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: 0609, 4560003  
; CURRENT APPLICATION NUMBER: US/10/623, 813  
; CURRENT FILING DATE: 2003-07-22  
; PRIOR APPLICATION NUMBER: US 09/720,066  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: PCT/US99/14373  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 60/090,906  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: US 60/093,993  
; PRIOR FILING DATE: 1998-07-24  
; NUMBER OF SEQ ID NOS: 119  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 4293  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-623-813-3  
Alignment Scores:  
Pred. No.: 0 Length: 4293  
Score: 3763.00 Matches: 689  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 9 Gaps: 0  
US-10-623-813-86 (1-689) x US-10-623-813-3 (1-4293)  
Qy 1 MetCAsnAlaValGluGlnGlnGlnGlyProGlyGluSerGlnLysValGluGlnAlaSer 20  
Db 899 ATGAATCTGTGGAAGAAACCAAGGGCCCGGAGTCTCAGAAAGTGAGAGAGCCAGC 958  
Qy 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu 40  
Db 959 CTTCTCTCTGTGACGACCCACTGACCCCTCCACTGCTGGCTACCAACCCCTGAG 1018  
Qy 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
Db 1019 CCGGTGGGTCCGATGCTGGGGAACAAGATGCCAAGAAGAGCGATGAGACCCAGAG 1078  
Qy 61 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTrrGlyLysLeuArgGlyPhe 80  
Db 1079 TACGAGACCGCCCGGCTTTGGCATTTGGGAGCTGTGTGGGGAACCTGCGGCTTC 1138  
Qy 81 SerTrpTrpProGlyArgGlyLeuValSerTrpTrpMetThrGlyArgSerArgAlaGlu 100  
Db 1139 TCTGTGTGGCCAGCCCATTTGTCTTGTGGATGACGCGCCGAGCCAGACGCTGAA 1198  
Qy 101 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValAlaCysValGluLys 120  
Db 1199 GGACCCGCTGGGTCAITGTGTTCGGAAGCGCAAAATTTCTCAGTGTGTGTGGAGAG 1258

Qy 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyraAsnLysGlnPro 140  
Db 1259 CTGATGCCGCTGACTGTTTGTGCAGTGCTTCCACAGGCCACGTAACACAGACAGCCC 1318  
Qy 141 MetTyraArgLysAlaIleTyrrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
Db 1319 ATGTACCGCAAGCATCTAGAGAGTCTCTGAGGTGGCCAGACCGCGGGGGAAGCTG 1378  
Qy 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluValAsn 180  
Db 1379 TTCCCGGTGTGCCACGACAGAGATGACACTCCCAAGGCCGTGGAGGTCCAGAAC 1438  
Qy 181 LysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
Db 1439 AAGCCATGATGTGATGGCCCTGGGGGGCTTCCAGCTTCTGCGCTTAAGGCTTGAG 1498  
Qy 201 ProProGluGluGluLysAsnProTyrrLysGluValTyrrThrAspMetTrpValGluPro 220  
Db 1499 CACCAAGAAAGAGAGATCCCTTCAAGAAGTGTACACGACATGTGGTGGAACCT 1558  
Qy 221 GluAlaAlaAlaTyrrAlaProProProAlaLysLysProArgLysSerThrAlaGlu 240  
Db 1559 GAGGCACTGCTTACGACACCACTCCACAGCCAAAGCCCGGAAGAGACAGCGGAG 1618  
Qy 241 LysProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrrGluVal 260  
Db 1619 AAGCCAAAGGTCAAGGAGATTTATGTAGCGCACAAAGAGCCGCTGTGTACAGAGTG 1678  
Qy 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
Db 1679 CCGGAGAAGTCCGGGAACATTGAGAGATCTGATCTCCTGTGGAGACCTCAATGTTACC 1728  
Qy 281 LeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
Db 1739 CTGGAAACACCCCTCTTGTGGAGATGTGCCAAATGCAAGAACTGCTTCTCGAG 1798  
Qy 301 CysAlaIleTyrrGlnTyrrAspAspAspGlyTyrrGlnSerTyrrCysThrIleCysCysGly 320  
Db 1799 TGTGCTTACCACTACGACGACGAGCGCTTACCACTCTCTGACCACTGCTGTGGGGGC 1858  
Qy 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
Db 1859 CGTGAGGTGCTCATGTGCGGAAACAACACTGCTGAGGTCTTTGCGTGAAGTGTG 1918  
Qy 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTrpAsnCys 360  
Db 1919 GACCTCTTGTGGGCGCGGGGCTGCCACAGCACTTAAGGAAGACCCCTGGAACCTGC 1978  
Qy 361 TyrMetCysGlyHisLysGlyTyrrTyrrGlyLeuLeuArgArgArgGluAspTrpProSer 380  
Db 1979 TACATGTGCGGACACAGGGTACTTACGAGCTGTGCGGCGGACAGAGACTGCCCTTC 2018  
Qy 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLysValTyr 400  
Db 2039 CCGGTCCAGATGTTCTTCGCTAATAACACAGACAGAAATTTGACCTCCAAAGTTTAC 2098  
Qy 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
Db 2099 CCACTGTCCACCTGGAAGAGAGAGACCCTCGGGTGTCTCTCTTTGAGGAGATC 2158  
Qy 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrrIleAlaSer 440  
Db 2159 GCTACAGGGCTCTCTGGTGTGAAGACTTGGCATTTAGGTGACCCCTACATTGCTCG 2218  
Qy 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 460  
Db 2219 GAGGTGTGTAGGATCTCATCAGGTGGCATGTGGGACACCGAGGAAGATCATGTAC 2278  
Qy 461 ValGlyAspValArgSerValThrGlnLysHisIleIleGlnLysTrpGlyProPheAspLeu 480  
Db 2279 GTGCGGAGCTCCGACGCGTCAACAGAAACATATCCAGAGAGTGGGGCCCATTCGATCTG 2338

```

Qy 481 ValIIEGIYGIYserProCYsaAspAspIleuSerIleValaenProAlaArgyluGlyIleu 500
Db 2339 GTGATGGGGGGAGTCCCTGGCAATGACCTCTCCATGCTCAACCTGCTGGCAAGGGCCCTC 2398
Qy 501 TyrgIuGIYThrGIYArgIleuPhePheGluPheTyrgIleuLeuHisAspAlaArgPro 520
Db 2399 TACGAGGGCACACTGGCGGCTCTTTTGAGTTCTACCCGCTCTCCAGATGATGCCGGGCC 2458
Qy 521 LyseGIuGIYAspAspArgProPhePheTrPleuPheGluAsnValaIAlaMetGIYVal 540
Db 2459 AAGGGGAGATGATGATGCCCTCTTCTTGCGCTCTTGTGAAGATGGTGCCATGGCCCTT 2518
Qy 541 SerAspLySaArgAspIleSerArgPheLeuGluIleuSerAsnProValMetIleAspAlaYs 560
Db 2519 AGTGAACAAGAGGACATCTCCGGAATTTCTCGAGTCAACCTCTGTATGATGATGCCAAA 2578
Qy 561 GluValSerAlaAlaHisArgAlaArgTygPheTrpGIYAsnLeuProGIYMeCAsnArg 580
Db 2579 GAAGTGTACGTCGACACAGGGCCCGCTACTTCTGGGGTAACTTCCCGGTATGAAACAG 2638
Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGluIleuGluGluCysLeuGluHisGIYArg 600
Db 2639 CCGTTGGCATCCACTGTGAATGATAGCTGAGCTGAGCTGAGAGTGTCTGAGAGCATGGCAG 2698
Qy 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGluGly 620
Db 2699 ATACCCAAAGTTCAGAAAGTGAAGACCAATCTACTACAGAGTCAAACTTCCATTAAGCAGG 2758
Qy 621 LysAspGlnHisPheProValPheMetAsnGluYsGIYAspIleLeuTrpCysThrGlu 640
Db 2759 AAAGACCAAGCATTTTCCCTGTCTTCAATGAATGAAGAAAGAGACATCTTATGGTGCACTGA 2818
Qy 641 MetGluArgValPheGlyPheProValHisTygThrAspValSerAsnMetSerArgLeu 660
Db 2819 ATGCAAAAGGATATTGGTTTCCAGTCCACTATGACGTCTCCAACTAGAGCGGCTTG 2878
Qy 661 AlaArgGlnArgLeuLeuGluArgSerTrpSerValProValIleArgHisLeuPheAla 680
Db 2879 GCGAGGCAAGACTGCTGGCCGGCTCATGAGCGGTGACGATCATCCGCACTTCCCT 2938
Qy 681 ProLeuYsGIYTygPheAlaCysVal 689
Db 2939 CCCGTGAAGAGTATTGTCGCGTGTG 2965

RESULT 8
US-10-623-813-83
; Sequence 83, Application US/10623813
; Publication No. US20040234997A1
; GENERAL INFORMATION:
; APPLICANT: Li, En
; APPLICANT: Okano, Masaki
; APPLICANT: Xie, Shaoping
; APPLICANT: Chen, Taiping
; TITLE OF INVENTION: De Novo DNA Cytosine Methyltransferase Genes, Polypeptides & Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0609.4560003
; CURRENT APPLICATION NUMBER: US/10/623.813
; CURRENT FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: US 09/720,086
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: PCT/US99/14373
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 60/090,906
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 60/093,993
; PRIOR FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 119
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 83
; LENGTH: 2318
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-623-813-83

```

```

Alignment Scores:
Pred. No.: 0 Length: 2318
Score: 3714.00 Matches: 679
Percent Similarity: 98.8% Conservative: 2
Best Local Similarity: 98.5% Mismatches: 8
Query Match: 98.7% Indels: 0
Gaps: 0
US-10-623-813-86 (1-689) x US-10-623-813-83 (1-2318)

Qy 1 MetAsnAlaValGIuGIYAsnGluGlyProGIYGlueSerGluYsAlGIuAlaSer 20
Db 148 ATGATGCTGTGGAAGAAGAACAGACCTCTGGAAGTCTCAAGAGTGAAGAGGAGCCAGC 207
Qy 21 ProProAlaValGIuGluInProThrAspProAlaSerProThrValAlaThrThProGIu 40
Db 208 CTTCTGCTGTGCGAGCCAGCCAGACCTGTCTCTCGACTGTGGCCACACCCCTGAG 267
Qy 41 ProValGIYSerAspAlaGIYAspLysAsnAlaThrLysAlaGIYAspAspGIUProGIu 60
Db 268 CCACTAGAGAGGGATGCTGGGACAAAGATGCTACCAAGCAGCCGAGATGAGCTGAG 327
Qy 61 TyrgIuAspGIYArgGIYpheGIYIleGIYGlueValITrpGIYLysLeuArgGIYphe 80
Db 328 TATGAGATGGCCGGGCTTTGGCATTGAGAGCTGTGTGGGAAACTTCGGGGCTTC 387
Qy 81 SerTrpTrpProGIYArgIleValSerTrpTrpMetThrGIYArgSerArgAlaAlaGIu 100
Db 388 TCCGTGTGGCCAGGCCCAATTTGTCTTGTGTGAATGACAGGCCGAGACCCGAGCTGAA 447
Qy 101 GlyThrArgTrpValMetTrpPheGIYAspGIYLysPheSerValaIAlaCysValGIYLys 120
Db 448 GGCACGTGCTGGGTCACTGTGTTGGAGATGCAAGATTCAGTGTGTGTGTGAGAG 507
Qy 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTygAsnLysGluPro 140
Db 508 CTCATGCCGCGTACACTCTCTCTGAGTGCATTCCACAGGCGCACTTCAACAAACAGGCC 567
Qy 141 MetTyrgIYAlaIleTyrgIuValLeuGluValaIleSerSerArgAlaGIYLysLeu 160
Db 568 ATGTACCGCAAGACCATCTAGAGTCTCCAGGTGCGACGACCCGTGGGGAAGCTG 627
Qy 161 PheProValCysHisAspSerAspLysSerAspThrAlaLysValaIAlaGIuAlaAsn 180
Db 628 TTTTCACTTGCCATGTGACATGATGAAGTGAACGTGCAAGCTGTGGAAGTCAAGAC 687
Qy 181 LysProMetIleGIuTrpAlaLeuGIYGIYpheGlnProSerGIYProLysGIYLeuGlu 200
Db 688 AAGCAGATGATGAATGGGCCCTCGGTGGCTTCCAGCCCTCGGCTTAAGGCCCTGAG 747
Qy 201 ProProGluGIuGIYLysAsnProTyrgIuValaIleTyrgIuValaIleThrAspMetTrpValGIuPro 220
Db 748 CCACCAAGAGAGGAAGATCTCTTACAGGAATTACACCGCAAGTGGTGGAGCT 807
Qy 221 GluAlaIAlaIAlaTyrgAlaProProProProAlaLysLysProArgLysSerThrAlaGIu 240
Db 808 GAAGCAGCTGCTTACGCCCAACCCCAAGCCCAAGAAACCCAGAAAGAGCACAACAGAG 867
Qy 241 LysProLysValLysGIuIleIleAspGIuArgThrArgIuArgLeuValaIleGIuVala 260
Db 868 AAACCTAAGGTCAAGAGATCATTTGATGAGGCAAGAGGCGCGTGTATGAGGTG 927
Qy 261 ArgGluLysCysArgAsnIleGIuAspIleCysIleSerCysGIYSerLeuAsnValaThr 280
Db 928 CGCCAGAGGTGCAAGAAACATGAGAGCATTTGTATCTCATGTGGAGCTTCAATGTCAAC 987
Qy 281 LeuGIuHisProLeuPheValaGIYGIYMetCysGIuAsnAsnLysLysAsnAsnCysPheLeuGIu 300
Db 988 CTGAGACCACTCACTTCTTATTTGAGGATGTGCGAGAACTGTAAAGACTGTTCTTGAG 1047
Qy 301 CysAlaTyrgIuTyrgAspAspAspGIYTyrgIuSerTyrgIuSerTyrgIuSerTyrgIuSer 320

```

Db 1048 TGTGCTTACCAAGTATGACGACGATGGTACCAGTCTCATTCGACCATCTGCTGGGGG 1107  
Qy 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 340  
Db 1108 CGTGAAGGCTCAAGTGTGGGAAACAACATGCTGCAAGTCTTTGTGTGAGGTGTG 1167  
Qy 341 AspleuLeuValGlyProGlyAlaAlaGlnAlaAlaIleValGluAspProTyrAsnGly 360  
Db 1168 GATCTCTTGGTGGGGCCAGAGAGCTGCTCAGGACCATTAAGAAAGACCTTGGAACTGC 1227  
Qy 361 TyrMetCysGlyHisGlyThrTyrGlyLeuLeuArgArgGluAspTyrProSer 380  
Db 1228 TACATGTGCGGCGCATTAAGGCGACCTATGGCTGCTGCACAAAGCGGAAGACTGGCTTCT 1287  
Qy 381 ArgLeuLinetPhePheAlaAsnAsnHisAspGlnGluPheAspProProValTyr 400  
Db 1288 CGACTCCAGATGTTCTTTGCAATTAACATGACCAAGCAATTGACCCCAAGGTTTAC 1347  
Qy 401 ProProValProAlaGluValArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
Db 1348 CCACCTGTGCCAGCTGAGAAAGAGAACCCCATCCCGCTGCTCTCTTGTGATGGGATT 1407  
Qy 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
Db 1408 GCTACAGGCGCTCCGTGCTGAAGGACCTGGGCACTCCAAAGTGACCCGCTACATGGCTCC 1467  
Qy 441 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyValIleMetTyr 460  
Db 1468 GAGGTGTGAGAGACTCATCACTGAGCGTGGGACCGGAGGAAAGATATATATAC 1527  
Qy 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyLysProPheAspLeu 480  
Db 1528 GTGGGGAGCTCCCGACGCTCAACAAGCAATTCAGAGAGTGGGGCCATTGACCTG 1587  
Qy 481 ValIleGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
Db 1588 GTGATTGAGAGGAGTCCCTCAATGACCTGCTCAATGCAACCTGCGCCCAAGGACTT 1647  
Qy 501 TyrGluGlyThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
Db 1648 TATGAGGGTACTGAGCGGCTCTTTGAGTTCACCGGCTCCGAGTATGCGGGGCC 1707  
Qy 521 LysGluGlyAspAspArgProPhePheTyrLeuPheGluAsnValValAlaMetGlyVal 540  
Db 1708 AAGGAGGAGATGATCGCCCTTCTTCTGCTCTTGAAGATGTGGTGGCCATGGGCTT 1767  
Qy 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
Db 1768 AGTGACAAAGAGGACATCTCGCATTTCTTGAGTCTAACCCCGTGAATGATGACGCCAA 1827  
Qy 561 GluValSerIleAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArg 580  
Db 1828 GAAGTGTGCTGCTCAACAAGGCGGCTTCTTCTGGGGTAACTTCTTGCAATGAACAGG 1887  
Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArg 600  
Db 1888 CTTTGGCATCCACTGGAATGATAGCTGAGAGCTGCAAGCTCTCGAGACAGCGAGA 1947  
Qy 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleValGlnGly 620  
Db 1948 ATAGCAAGTTCAAGAAAGAGGACCATTAACCAAGGTCAAACTCTATAAAGCAGGGC 2007  
Qy 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGlu 640  
Db 2008 AAAGACCAAGATTTCCCGCTTCATGAACGAAAGAGGAGCAATCTGTGTGATGAGAA 2067  
Qy 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
Db 2068 ATGGAAGAGGTGTGTTGGCTTCCCGTCACTACACAGACGCTTCCAAAGAGCGGCTTG 2127  
Qy 661 AlaArgGlnArgLeuGluArgSerTyrSerValProValIleArgHisLeuPheAla 680  
Db 2128 GCGAGGACAGAGTGTGCGGCGCATGTGAGCGTGGCGGTCACTCCGCACTTCTGCT 2187

Qy 681 ProLeuLysGluTyrPheAlaCysVal 689  
Db 2188 CCGCTGAAGGAATATTTGCTGTGTG 2214

RESULT 9  
US-09-720-086-1  
; Sequence 1, Application US/09720086  
; Publication No. US20060084053A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, En  
; APPLICANT: Okano, Masaki  
; APPLICANT: Xie, Shaoqing  
; TITLE OF INVENTION: De Novo DNA Cytosine Methyltransferase Genes,  
; FILE REFERENCE: 0609,4560002  
; CURRENT APPLICATION NUMBER: US/09/720,086  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR FILING DATE: PCT/US99/14373  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/090,906  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/093,993  
; PRIOR FILING DATE: 1998-07-24  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4192  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: Unsure  
; LOCATION: (4161)..(4161)  
; OTHER INFORMATION: May be any nucleic acid  
; US-09-720-086-1

Alignment Scores:  
Score: 0 Length: 4192  
Percent Similarity: 3714.00 Matches: 679  
Best Local Similarity: 98.8% Conservative: 2  
Query Match: 98.5% Mismatches: 8  
DB: 98.7% Indels: 0  
Gaps: 0

US-10-623-813-86 (1-689) x US-09-720-086-1 (1-4192)

Qy 1 MetAsnAlaValGluGluLeuGlnGlyProGlyGluSerGlnLysValGluGluAlaSer 20  
Db 874 ATGAATGCTGTGGAAGAGAAACAGGCTCTGAGAGATCTCAGAAAGTGGAGAGCCAGC 933  
Qy 21 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrTyrProGlu 40  
Db 934 CTTCTGCTGTGTGAGAGAGCCAGGACCTGCTTCCGACTGTGGCCACACCCCTGAG 993  
Qy 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
Db 994 CCAAGTGAAGGAGATCTGGGAGCAAGATCTCAACCAAGCACCGACATGAGCTGAG 1053  
Qy 61 TyrGluAspGlyValArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPhe 80  
Db 1054 TATGAGAGATGGCGGGGCTTTGGCATTTGAGAGAGCTGTGTGGGGAAACTTCCGGGCTTC 1113  
Qy 81 SerTyrTyrProGlyValArgIleValSerTyrPheMetThrGlyValArgSerArgAlaGlu 100  
Db 1114 TCTGTGTGGCCAGGCGGAAATGTGTCTTGTGGATGACAGGCGGAGCGAGCGCTGAA 1173  
Qy 101 GlyThrArgTyrValMetTyrPheGlyAspGlyLysPheSerValValCysValGluLys 120  
Db 1174 GCGACCTGGCGGTATGTGTGTTGCGAGATGCGCAATTTCTCACTGTGTGTGTGAGAAAG 1233  
Qy 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140  
Db 1234 CTCATCCCGCTGAGCTCTTCTGCAATGATTCACACAGGCCCACTTAACAAGCAGCCC 1293

QY 141 MetYrAglYsAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
 Db ATGTACCCCAAGCATCTACCAAGTCTCTCCAGGTGGCCAGCAGCCGTGCCGGGAAGCTG 1353  
 QY 161 PheProValCybHisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsn 180  
 Db TTTCACGCTTGCACGACGATGATGAAGTACAGTACAGGCTGTGGAAGTGCAGAAC 1413  
 QY 181 LysProMetIleGluTrpAlaLeuGlyIlePheGlnProSerGlyProLysGlyLeuGlu 200  
 Db AAGCAGATGATTGAATGGGCCCTCGGTGGCTTCCAGCCCTGGGTCCTTAAGGCCCTGAG 1473  
 QY 201 ProProGluGluGluLysAsnProTyrIleGluValIleTyrAspMetTyrValGluPro 220  
 Db CCACCAAGAAAGAAAGAAATCTTACAGAGAGTTTACCCGACATGTGGTGGAGCTT 1533  
 QY 221 GluAlaAlaIleTyrAlaProProProAlaLysLysProArgLysSerThrAlaGlu 240  
 Db GAAGCAGCTGCTTACGCCCCCACCACCGCAAGAAACCAGAAAGACACACAGAG 1593  
 QY 241 LysProLysValLysGluIleIleAspGluArgThrArgLysArgLysValTyrGluVal 260  
 Db AAACCTAAGCTCAAGAGATCATTCATGAGGCAACAGAGAGCGGTGTGTATGAGCTG 1653  
 QY 261 ArgGlnLysCybAspAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
 Db CGCCAGAAAGTCAGAAACATCGAGACATTTGTATCTCATGTGGAGCCTCAATGTCC 1713  
 QY 281 LeuGlnHisProLeuPheValGlyIleMetCysGlnAsnGlyLysAsnCybPheLeuGlu 300  
 Db CTGGAGCACCACCTTCATTTGAGGACATGTGCCAACTGTMAACCTGCTCTTGAG 1773  
 QY 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGly 320  
 Db TGTGCTTACCAATGACGACGATGATGATGATGATGATGATGATGATGATGATGATG 1833  
 QY 321 ArgGluValLeuMetCysGlyLysAsnAsnGlyCysAspGlyCysPheCysValGluCysVal 340  
 Db CGTGAAGTCTCATGTGTGGGAAACAACATCTGCAGAGTGTGTGTGTGTGTGTGTGT 1893  
 QY 341 AspleuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyrAsnCys 360  
 Db GATCTCTTGTGGGCCAGAGCTCTCTAGGAGCATTAAGAGACCCCTGGAAGCTGC 1953  
 QY 361 TyrMetCysGlyHisLysGlyThrTyrIleLeuLeuArgArgGluAspTyrProSer 380  
 Db TACATGTGCGGACATTAAGGACCTATGCGCTGTGCGAAGACGGGAGACTGGCTTCT 2013  
 QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 400  
 Db CGACTCCAGATGTCTTGTGCAATTAACATGACCGAATTGTGACCCCAAGAGTTTAC 2073  
 QY 401 ProProValProAlaGluLysArgLysProIleArgValLysSerLeuPheAspGlyIle 420  
 Db CCACCTGTGCCAGCTGAGAAAGAGAGCCCATCCGCGTGTCTCTCTTTTGAAGGATT 2133  
 QY 421 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAspSer 440  
 Db GCTACAGGGGCTCTGTGTCTGAGAGACTGGGACATCAAGTGAACCGCTACATTCCTCC 2153  
 QY 441 GluValCybGluAspSerIleThrValIleMetValAlaGlnIleGlnGlyLysIleMetTyr 460  
 Db GAGGTGTGTGAGACTCATACCGTGGCATGTGTGGCCACCAAGAAAGATCATGTGAC 2253  
 QY 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeu 480  
 Db CTCGGGAGACGTCCGAGGCTCACACAGACATATCCAGAGTGGGGCCATTTGACCTG 2313  
 QY 481 ValIleGlyLysProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
 Db GTGATTGAGGACAGTCCCTGCAATGACCTTCATGTCACCCCTGCCGAGGAGGACTT 2373

QY 501 TyrGluGluTyrThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAspAlaArgPro 520  
 Db TATAGGGTACTGGCCCTCTCTTTGAGTTCTACCGCTCTGCAATGATGCCGGGCC 2433  
 QY 521 LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnValAlaIleMetGlyVal 540  
 Db AAGAGGGAGATGATCCCTCTTCTTGTGCTCTTTAGAAATGTGTGGCATGGGCTT 2493  
 QY 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
 Db AGTACCAAGAGGGACATCTCGCATTTCTTGAGTCTTAACCCCGATGATGATGAGCAAA 2553  
 QY 561 GluValSerAlaIleHisArgAlaArgTyrPheTrpGlyLysLeuProGlyMetAsnArg 580  
 Db GAAGTGTGTGTGACACAGAGGCCCGTTACTTCTGGGGTAACTTCTTGGCATTAACAG 2613  
 QY 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGlyCysLeuGlnHisGlyArg 600  
 Db CTTTGGCATCTCACTGGAATGATTAAGCTGAGCTGCAGAGTGTCTGAGCAGCGCAGA 2673  
 QY 601 IleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGly 620  
 Db ATAAGCAAGTTCAAGCAAGAGAGACCATTAACCAAGGTCAACTGTATTAACAGAGGC 2733  
 QY 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGlu 640  
 Db AAAGACCAAGCTTTCCTCCGCTTCTATGACAGAGAGAGAGACATCTGTGTGATGAA 2793  
 QY 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
 Db ATGGAAGGGGTGTGTGCTTCCCGCTCACTACACAGACGCTCCAAATGAGCGGCTTG 2853  
 QY 661 AlaArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAla 680  
 Db GCGAGCGAGAGACTGCTGGGCGCATCTGTGAGAGCGTGGCTCATCCGCACTTCTGCT 2913  
 QY 681 ProLeuLysGluTyrPheAlaCysVal 689  
 Db CGGCTGAAGGAATTTTGTGCTGTGTG 2940

Db 2914 CGGCTGAAGGAATTTTGTGCTGTGTG 2940

RESULT 10  
 US-10-623-813-1  
 : Sequence 1, Application US/10623813  
 : Publication No. US20040234997A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Li, En  
 : APPLICANT: Okano, Masaki  
 : APPLICANT: Xie, Shaoping  
 : APPLICANT: Chen, Taiping  
 : TITLE OF INVENTION: De Novo DNA Cytosine Methyltransferase Genes, Polypeptides & Uses  
 : TITLE OF INVENTION: Theaof  
 : FILE REFERENCE: 0609.4560003  
 : CURRENT APPLICATION NUMBER: US/10/623.813  
 : PRIOR FILING DATE: 2003-07-22  
 : PRIOR APPLICATION NUMBER: US 09/720,086  
 : PRIOR FILING DATE: 1999-06-25  
 : PRIOR APPLICATION NUMBER: PCT/US99/14373  
 : PRIOR FILING DATE: 1999-06-25  
 : PRIOR APPLICATION NUMBER: US 60/090,906  
 : PRIOR FILING DATE: 1998-06-25  
 : PRIOR APPLICATION NUMBER: US 60/093,993  
 : PRIOR FILING DATE: 1998-07-24  
 : NUMBER OF SEQ ID NOS: 119  
 : SOFTWARE: PatentIn version 3.2  
 : SEQ ID NO 1  
 : LENGTH: 4192  
 : TYPE: DNA  
 : ORGANISM: Mus musculus  
 : FEATURE:  
 : NAME/KEY: misc\_feature  
 : LOCATION: (4161)..(4161)  
 : OTHER INFORMATION: n is a, c, g, or t  
 US-10-623-813-1

## Alignment Scores:

Pred. No.:	0	Length:	4192
Score:	3714.00	Matches:	679
Percent Similarity:	98.8%	Conservative:	2
Best Local Similarity:	98.5%	Mismatches:	8
Query Match:	98.7%	Indels:	0
DB:	9	Gaps:	0

US-10-623-813-86 (1-689) x US-10-623-813-1 (1-4192)

Qy	1	Metasna1aVal1GluGluasnGlnGlyProGlyGlySerGlnLysVal1GluGlu1a1aSer	20
Db	874	ATGATGCTGTGGAAGAACACAGGCCCTGTGAGAGTCTCAAGAGTGTGAGAGGCCACG	933
Qy	21	ProPro1a1aVal1GlnGlnProThrAspPro1a1aSerProThrVal1a1aThnThProGlu	40
Db	934	CTCTCTGCTGTGCGACAGCCACCGACCTCTCTCCGACTGTGGCCACCACTTGTAG	993
Qy	41	ProValGlySerAsp1a1aGlyAspLysAsn1a1aThnLys1a1aGlyAspAspGluProGlu	60
Db	994	CCAGTAGAGGGAGTGTGGGGACAGATCTACCAAGACGCCGAGATGAGCTGAG	1053
Qy	61	TyrGluAspGlyArgGlyPheGly11eGlyGluLeuVal1TrpGlyLysLeuArgGlyPhe	80
Db	1054	TATGAGATGGCCGGGCTTTGGCATTGAGAGCTGTGTGGGGAAACTTCGGGGCTTC	1113
Qy	81	SetTrpTrpProGlyArgGlyLevalSerTrpTrpMetThrGlyArgSerArg1a1aGlu	100
Db	1114	TCTCGTGGCCAGGCCCAATGTGTCTTGTGTGATGACAGGCCGGAAGCCGACAGCTGAA	1173
Qy	101	GlyThrArgTrpVal1MetTrpPheGlyAspGlyLysPheSerVal1a1aCysVal1GluLys	120
Db	1174	GGCACTGCTGGGTCATGTGTGTGAGATGCAAGTCTCAGTGTGTGTGTGAGAG	1233
Qy	121	LeuMetProLeuSerSerPhePheCysSer1a1aPheHisGln1a1aThnTyrAsnLysGlnPro	140
Db	1234	CTCATGCGGTGAGCTCTTCTGTGAGTGCATTCACACGAGCCACTTCACAAACAGAGCCC	1293
Qy	141	MetTyrArgLysAla1eTyrGluVal1LeuGlnVal1a1aSerSerArg1a1aGlyLysLeu	160
Db	1294	ATGTACCGCAAGCATCTACGAGTCTCTCCAGGTGGCCAGACCCGTCGGGAGAGCTG	1353
Qy	161	PheProVal1a1aLysHisAspSerAspGlySerAspThr1a1aLysVal1a1aGluVal1a1a	180
Db	1354	TTTCCAGCTTGCATGACAGATGAAGTACAGTGGCAAGCTGTGAGAGTGCAGAAC	1413
Qy	181	LysProMet11eGlnTrp1a1aLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu	200
Db	1414	AAGCAGATGATGATGGCCCTCGTGGCTTCCAGCCCTCGGCTCCTAAGGGCTGGAG	1473
Qy	201	ProProGluGluGlyLysAsnProTyrLysGluVal1a1aThrAspMetTrpVal1GluPro	220
Db	1474	CCACCAAGATGATGATGGCCCTCGTGGCTTCCAGCCCTCGGCTCCTAAGGGCTGGAG	1533
Qy	221	GluAla1a1a1a1aTyr1a1aProProProPro1a1aLysLysProArgLysSerThn1a1aGlu	240
Db	1534	GAAAGACCTGTTAGCCGCCACCCACACAGCCCAAGAAACCCAGAAAGAGCAACAGAG	1593
Qy	241	LysProLysVal1LysGlu1e11eAspGlyLysArgThrArgGluArgLeuVal1a1aGluVal1	260
Db	1594	AAACCTAAGGTCAAGAGATCATTTGATGAGGCAACAGGGAGCGGCTGTATAGAGGTG	1653
Qy	261	ArgGlnLysCysArgAsn11eGluAsp11eCys11eSerCysGlySerLeuAsnVal1Thr	280
Db	1654	CGCCAGAAAGTCAAGAAACATCTGAGAGACATTTGTATCTCATGTGGAGACCTCAAGTCA	1713
Qy	281	LeuGlnHisProLeuPheVal1GlyLysCysGlnLysCysLysAsnCysPheLeuGlu	300
Db	1714	CTGAGAGACCACTTCTTATTTGAGGATGTGCCAGAACTGTAAAGAACTGCTTGTGAG	1773
Qy	301	CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThn11eCysCysGlyGly	320

Db	1774	TGTCTTACCAAGTATGACGAGATGGGTATGACCATCTATTGACCAATCTGCTGGGGG	1833
Qy	321	ArgGluVal1LeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysVal1GluCysVal1	340
Db	1834	CGTAAATGCTCATGTGTGTGGGAACAACATCTGTGAGAGTCTTTGTGTGAGTGTGTG	1893
Qy	341	AspLeuLeuVal1GlyProGlyVal1a1aGln1a1a1eLysGluAspProTrpAsnCys	360
Db	1894	GATCTCTTGGTGGGGCCAGAGAGCTGTCTCAGGCACTTAAAGAAAGCCCTGGAATGCG	1953
Qy	361	TyrMetCysGlyLysLysGlyThnTyrGlyLeuLeuArgArgArgGluAspTrpProSer	380
Db	1954	TACATGTGCGGGCATTAAGGACCACTTATGGCTGTGTGGAAGACCGGAAGACTGCTCT	2013
Qy	381	ArgLeuGlnMetPhePhe1a1aAsnAsnHisAspGlnGluPheAspProLysVal1Tyr	400
Db	2014	CGACTCCAGATGTTCTTGTGCAATTAACATGACCAAGAAATTTGACCCCCCAAGATTAC	2073
Qy	401	ProProVal1Pro1a1aGlyLysArgLysPro11eArgVal1LeuSerLeuPheAspGly11e	420
Db	2074	CCACCTTCCCACTGAGAAAGAGAACCCATCCGCTGTCTCTCTTGTGATGGATTT	2133
Qy	421	AlaThrGlyLeuLeuVal1LeuLysAspLeuGly11eGlnVal1a1aSerArgTyr11eAlaSer	440
Db	2134	GCTACAGGCTCTGCTGTGAAGGACCTGGGCATCCAGTGCACCGCTACATTGCCCTCC	2193
Qy	441	GluVal1CysGluAspSer11eThnVal1GlyMetVal1ArgHisGlnGlyLys11eMetTyr	460
Db	2194	GAGGTGTGTAGGACTCATACCGTGGCATGTGTGGCCACCGGAAGAAATATATGAC	2253
Qy	461	ValGlyAspVal1ArgSerVal1ThnGlnLysHis11eGlnGluTrpGlyProPheAspLeu	480
Db	2254	GTCGGGAGCTCCCGACCGTCACACAGAACATTCACAGAGTGGGCCCATTCGACCTG	2313
Qy	481	Val11eGlyLysSerProCysAsnAspLeuSer11eVal1a1aPro1a1aArgLysGlyLeu	500
Db	2314	GTGATGTGAGAGGACCTCTGCAATGACCTCTCCATTGTCAACCTGCGCCGCAAGGACTT	2373
Qy	501	TyrGlnGlyThrGlyArgLeuPhePheGluPheTyrArgLeuHisAspAlaArgPro	520
Db	2374	TATGAGGTACTGCGCCCTCTTGTGAAGTTCACCGCTCTCGACATGATGCCGGGCC	2433
Qy	521	LysGluGlyAspAspArgProPhePheTrpLeuPheGluAsnVal1a1a1aMetGlyVal1	540
Db	2434	AAGAGGGAATGATGCCCTCTTCTGTGCTCTTGAAGATGTGTGGCCATGGGCTT	2493
Qy	541	SerAspLysArgAsp11eSerArgPheLeuGluSerAsnProVal1Met11eAspAlaLys	560
Db	2494	AGTGACAAAGAGGACATCTGCGATTTCTTGAGCTTAACCCCGTGAATGATGAGCCAAA	2553
Qy	561	GluVal1Ser1a1aHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArg	580
Db	2554	GAAAGTCTCTGCACACAGGGCCCTTACTTCTGGGGTAACTTCTGGGATTAACAGG	2613
Qy	581	ProLeu1a1aSerThrVal1a1aAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArg	600
Db	2614	CTTTGGCATCCACTGTGAATGATTAAGCTGAGCTGCAAGAGTGTCTGAGACAGGCGAGA	2673
Qy	601	11eAlaLysPheSerLysVal1ArgThn11eThnThnArgSerAsnSer11eLysGlnGly	620
Db	2674	ATAGCCAAAGTTCCAGAAAGTGAAGACCATTAACCAACCGGTCAAACTGTATTAAGCAG	2733
Qy	621	LysAspGlnHisPheProVal1PheMetAspGluLysGluAsp11eLeuTrpCysThnGlu	640
Db	2734	AAAGACCAAGATTTCCCGTCTTATGAACAGAAAGAGGACATCTGTGTGTGACGTGA	2793
Qy	641	MetGluArgVal1PheGlyPheProVal1HisTyrThnAspVal1SerAsnMetSerArgLeu	660
Db	2794	ATGGAAGGGGTGTTGGCTTCCCGTCCACTACACACACATCTTCAACATGAGCGGCTTG	2853
Qy	661	AlaArgGlnArgLeuLeuGlyArgSerTrpSerVal1ProVal11eArgHis11ePhePheAla	680
Db	2854	GCGAGGACAGACTGCTGGGCGATGTGAGCGTGGCTGTATCCGCACTCTTGGCT	2913

QY 681 ProteulysGluTyrPheAlaCysVal 689  
DB 2914 CCGCTGAAGAAATATTTGCTGTGTG 2940

RESULT 11  
US-10-330-773-276  
; Sequence 276 Application US/10330773  
; Publication No. US20060040262A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001300  
; CURRENT FILING DATE: 2002-12-27  
; NUMBER OF SEQ ID NOS: 981  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 276  
; LENGTH: 6157  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-330-773-276

Alignment Scores:  
Pred. No.: 0 Length: 6157  
Score: 3714.00 Matches: 679  
Percent Similarity: 98.8% Conservative: 2  
Best Local Similarity: 98.5% Mismatches: 8  
Query Match: 98.7% Indels: 0  
Gaps: 0

US-10-623-813-86 (1-689) x US-10-330-773-276 (1-6157)

QY 1 MetaAlaValAlaGluAlaGlnGlnGlyProGlyGlnSerGlnLysValAlaGluAlaSer 20  
DB 866 ATGAATGCTGTGAAGAAACACGAGCCTCTGGAGAGTCTCAAGAGTGGAGAGCCAGC 925

QY 21 ProProAlaValAlaGlnGlnProThrAspProAlaSerProThrValAlaThrThProGlu 40  
DB 926 CCTCTGCTGTGACAGACCCACGAGCCTCTCTCGACTGTGGCCACCACTCCCTGAG 985

QY 41 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 60  
DB 986 CCAAGTGAAGAGGAGTGTGTGGGACAAAGATCTACCAAGACGAGAGAGAGAGCTGAG 1045

QY 61 TyrGluAspGlyArgGlyPheGlyTleGlyGluLeuValTyrGlyLysLeuArgGlyPhe 80  
DB 1046 TATGAGGATGGCCGGGCTTTGGCATTTGAGAGCTGTGTGGGGAACCTTCGGGGCTTC 1105

QY 81 SerTTPTPProGlyArgGlyLeuAlaSerTTPTPMetThrGlyArgSerArgAlaAlaGlu 100  
DB 1106 TCCGTGTGGCCAGGCGCAATTTGTCTTGTGTGAAGACAGGCGGAGCGGAGCTGAA 1165

QY 101 GlyThrArgTTPValMetTTPPheGlyLysPheSerValValCysValAlaGluLys 120  
DB 1166 GGCACCTGCTGGGTCATGTGTTCGAGATGGCAAGTTCTCAGTGTGTGTGTGAGAG 1225

QY 121 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 140  
DB 1226 CTCATGCCGCTGAGCTCTTCTGCGAGTCTCCACGAGCCACTTACAAAGAGAGCCC 1285

QY 141 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 160  
DB 1286 ATGTAACGCAAGCATCTACGAAGTCTCCAGGTGGCCAGAGCGCTGGGAGAGCTG 1345

QY 161 PheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValAlaGlnAsn 180  
DB 1346 TTTTCAAGCTTCCATGACAGTGAATAAGTGAACATGCGCAAGCGCTGGAAGTGCAGAC 1405

QY 181 LysProMetIleGlyTTPAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 200  
DB 1406 AAGCAGATGATTGAATGGGCTTCGTTCCAGCTTCGGGCTCTTAAGGCGCTTGAG 1465

QY 201 ProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTTPValGluPro 220  
DB 1466 CCACAGAAAGAGAAATCTTACAGAGAACTTTACCCAGCATGTGGTGAAGCTT 1525

QY 221 GluAlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSerThrAlaGlu 240  
DB 1526 GAACACAGCTGCTTACGCCGCCACCCACAGCCAAAGAACCCAGAAAGAGACAAACAGAG 1585

QY 241 LysProLysValLysGluIleLysPheLysArgThrArgGluValTyrGluVal 260  
DB 1586 AAACCTAAGGCAAGGAGATCATGTAGAGCGCAAGGAGCGCTGTGTAGAGAGTG 1645

QY 261 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 280  
DB 1646 GCGCAAGTGCAGAAACATCGAGAGCATTTGTATCTCATGTGGACCTCAATGTCAAC 1705

QY 281 LeuGluHisProLeuPheValGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 300  
DB 1706 CTGGAGACCACTCTTCAATGGAGCATGTGCCAAGACTGTAAAGAACTGCTTCTTGAG 1765

QY 301 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 320  
DB 1766 TGTGCTTACAGATGACGAGAGGTGTACAGTCTTATTCACCATCTGCTGTGGGGG 1825

QY 321 ArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValAlaCysVal 340  
DB 1826 CGTGAATGCTCATGTGTGGGAAACAACTGCTGAGAGTCTTTGTGTGATGTGTG 1885

QY 341 AspLeuLeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspProTTPAsnCys 360  
DB 1886 GATCTCTTGGTGGGCGAGAGCTGCTCAGGACCACTTAAAGAAAGCCCTCGAAGCTGC 1945

QY 361 TyrMetCysGlyHisLysGlyTyrTyrGlyLeuLeuArgArgArgGluAspTTPProSer 380  
DB 1946 TACATGTGCGGCAATGAAGGACCTATGAGCTGTGTGAAGACGGGAAGACTGGCTTCT 2005

QY 381 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLysValTyr 400  
DB 2006 GCACTCCAGATGTTCTTTGCAATTAACATGACAGGAATTTGACCCCCAAAGGTTTAC 2065

QY 401 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 420  
DB 2066 CCACCTGTCCAGCTGAGTGAAGAGAGAGCCATCGCTGTGTCTCTTTGATGGGATTT 2125

QY 421 AlaThrGlyLeuLeuValLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 440  
DB 2126 GCTACAGGCTCTCTGTGTGAAGAGACTGGGCACTCCAAAGTGAACCTTACATTGCCCTCC 2185

QY 441 GluValCysGluAspSerIleThrValAlaMetValArgHisGlnGlyLysIleMetTyr 460  
DB 2186 GAGGTGTGAGACTCTCATACAGGTGGCATGTGTGGGACCAAGGAAAGATATATATAC 2245

QY 461 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTTPGlyProPheAspLeu 480  
DB 2246 GTGCGGAGACGTCGACAGCTCACAGAAACATATCCAGAGTGGGCGCCATTGACCTTG 2305

QY 481 ValIleGlyLysProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 500  
DB 2306 GTGATTTGAAGAGCTCTCTCAATGACCTTCAATTTCAACCTTGGCCGGAAGGACTT 2365

QY 501 TyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuHisAspAlaArgPro 520  
DB 2366 TATACGGGTACTGGCCGCTCTTCTTGAATTTCTACCGGCTCTCGCATGATGGCGGCC 2425

QY 521 LysGluGlyAspAspArgProPhePheTTPLeuPheGluAsnValAlaIleMetGlyVal 540  
DB 2426 AAGGAGGAGATGATCCCTCTTCTGTGCTCTTTGAGATGTGGTGGCCATGGCGGCTT 2485

QY 541 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 560  
DB 2486 AGTACAAAGAGGACATCTGCGCAATTTCTTGAATTAACCCCGTATGATTGACGCCAAA 2545

Qy 561 GluValSerAlaAlaHisArgAlaArgTyrPheTrpGluAsnLeuProGluMetAsnArg 580  
Db 2546 GAAGTGTCTGTGTGACACAGGCCCCGTTACTTCTGGGTAACCTTCTGGCATGACACAG 2605  
Qy 581 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArg 600  
Db 2606 CTTTGGCATCTCACTGTAATGATAGCTGGAGCTGCAGAGTGTCTGGACACAGGCGAGA 2665  
Qy 601 IleAlaLysPheSerLysValArgThrIlePheThrArgSerAsnSerIleLysGlnGly 620  
Db 2666 ATAGCCAAAGTTCCACCAAGTAGAGCACTTACCAACAGTCAAACTCTATTAACAGAGGC 2725  
Qy 621 LysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGlu 640  
Db 2726 AAGACCAAGCATTTCCCGCTTTCATGACAGAAAGAGGACATCTGTGTGATGACATGAA 2785  
Qy 641 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 660  
Db 2786 ATGGAAAGGGGTGTGGCTTCCCGTCCCTCAACACAGACGTCTCAACATGAGCCGCTTG 2845  
Qy 661 AlaArgGlnArgLeuGluArgSerTrpSerValProValIleArgHisLeuPheAla 680  
Db 2846 GCGGCGAGAGACTGCTGGGCGCATCGTGAGCGTGCCTGATCCGCCACTTCTGCT 2905  
Qy 681 ProLeuLysGluTyrPheAlaCysVal 689  
Db 2906 CCGCTGAAGGAATATTTTCTTGTGTG 2932

RESULT 12  
US-10-264-237-953  
Sequence 953, Application US/10264237  
Publication No. US20040009491A1  
GENERAL INFORMATION:  
APPLICANT: Btse et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: P4131P1  
CURRENT APPLICATION NUMBER: US/10/264,237  
CURRENT FILING DATE: 2002-10-04  
PRIOR APPLICATION NUMBER: PCT/US01/16450  
PRIOR FILING DATE: 2001-05-18  
PRIOR APPLICATION NUMBER: US 60/205,515  
PRIOR FILING DATE: 2000-05-19  
NUMBER OF SEQ ID NOS: 2876  
SOFTWARE: PatentIn Ver. 3.1  
SEQ ID NO 953  
LENGTH: 2938  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (491)..(491)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1987)..(1987)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1994)..(1994)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (2892)..(2892)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (2932)..(2932)  
OTHER INFORMATION: n equals a,t,g, or c  
US-10-264-237-953

Alignment Scores:  
Pred. No.: 2,11e-293 Length: 2938  
Score: 2713.00 Matches: 496

Percent Similarity: 99.0% Conservative: 0  
Best Local Similarity: 99.0% Mismatches: 5  
Query Match: 72.1% Indels: 0  
DB: 7 Gaps: 0

US-10-623-813-86 (1-689) x US-10-264-237-953 (1-2938)

Qy 189 GlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGluGluLysAsnPro 208  
Db 1 GGCACGAGCCAGGCTTTGTGGCCCTTAAGGCGCTTAGAGCCACACAGAAAGAAATCC 60  
Qy 209 TyrLysGluValTyrThrAspMetTrpValGluProGluAlaAlaTyrAlaProPro 228  
Db 61 TACAAAGATGTATACACGACATGTGGGTGAACCTGAGCAGCTGCTACGACACACT 120  
Qy 229 ProProAlaLysLysProAlaGlySerThrAlaGluLysProLysValLysGluIle 248  
Db 121 CCACAGCCCAAAAGCCCGGAAAGACACAGCGAGAGGCCAAGTCAAGTCAAGAGATTATT 180  
Qy 249 AspGluArgThrArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGlu 268  
Db 181 GATGAGGCGCAAGAGAGCGGCTGTGTACGAGTGGCGAGAAAGTCCGGAACATTGAG 240  
Qy 269 AspIleCysIleSerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGly 288  
Db 241 GACATCTGATCTCTCTGTGGAGCTTCATGTTACCTGGAAACCCCTCTTGTGTGA 300  
Qy 289 GlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysValTyrGlnTyrAspAsp 308  
Db 301 GGAATGTCCAAACCTGCAGAACTGCTTGTGAGTGTGACCTACCTACGACGACGAC 360  
Qy 309 GlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsn 328  
Db 361 GGCTACAGTCTTACTGACCATCTGTGTGGGCGCTGAGGTGCTCATGTGGGAAC 420  
Qy 329 AsnAsnCysCysArgCysPheCysValGluCysValAspLeuValGlyProGlyAla 348  
Db 421 AACACTGTGACGAGTCTTTCGTGAGTGTGACCTCTTGGTGGGCGGCGGCT 480  
Qy 349 AlaGlnAlaAlaLysGluAspProTyrAsnCysIleTyrMetCysGlyHisLysGlyThr 368  
Db 481 GSCCARCAGNCATTTAAGAAAGACCCCTGGAACCTGCTACACTGTGGCGGCAAGGATACC 540  
Qy 369 TyrGlyLeuLeuArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsn 388  
Db 541 TACGGGTGCTGGGCGCGGAGAGACTGSCCTCCGGTCCAGATGTTCTTGGTAAT 600  
Qy 389 AsnHisAspGlnLysPheAspProProLysValTyrProProValProAlaGluLysArg 408  
Db 601 AACACGACCAAGAAATTTGACCTCCAAAGTTTACCACCTGCCCAGCTGAAGAGG 660  
Qy 409 LysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLys 428  
Db 661 AAGCCCATCCGGGTGCTCTCTTTGATGGAATCGTACAGGCTCTGTGTCTGAAG 720  
Qy 429 AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThr 448  
Db 721 GACTTGGGCAATTAGGTGACCCGCTACATTGCTCGGARGTGTGTGACCTCATACAG 780  
Qy 449 ValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThr 468  
Db 781 GTGGGATGTGTGGCGCACCGGGGAGATCATGTACCTCGGAGACGTCACAGCTGACA 840  
Qy 469 GlnLysHisIleGlnGluTyrProPheAspLeuValIleGlyGlySerProCysAsn 488  
Db 841 CAGAAGCATATCCAGAGAGTGGGCCCATTCGATGTGTGTGGGGCAGTCCCTGCAT 900  
Qy 489 AspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhe 508  
Db 901 GACTCTCCATCTGTAACCTGTCTGGAAGGCTCTACAGAGGCACTGGCCGCTCTTC 960  
Qy 509 PheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhe 528

```

Db      961 TTTGAGTCTTACCGGCTCTGCATGATGCGCGCCCAAGAGGAGATGATGCCCTTC 1020
Oy      529 PheTTPLeuPheGluAsnValValAlaMetGlyValSerAspLysArgAspIleSerArg 548
Db      1021 TTTCTGGCTCTTTGAAATGCTGGTGGCCATGGGCGTTAGTACAAAGAGGACATCTCGCGA 1080
Oy      549 PheLeuGluSerAspProValMetIleAspAlaLysGluValSerAlaAlaHisArgAla 568
Db      1081 TTTCTCGAGTCCAACTCTGTGATGATGATGTCCAAAGAGTGTCTGCTGACACACAGGGCC 1140
Oy      569 ArgTTPheTTPGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAsp 588
Db      1141 CGCTACTTCTGGCGGTAACTTCCCGTATGACACAGCGCTTGTCATCCATCTGTGATGAT 1200
Oy      589 LysLeuGluLeuGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArg 608
Db      1201 AAGCTGGAGCTGCAGAGAGTGTCTGAGCATGCGACAGATAGCCAACTTCACGAAAGTGAAG 1260
Oy      609 ThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPhe 628
Db      1261 ACCATTACTACGAGGTCAAACCTCCATPAAAGCAGGSCAAAGACACACATTTTCTCTTC 1320
Oy      629 MetAsnGluLysGluAspIleLeuTTPCysThrGluMetGluArgValPheGlyPhePro 648
Db      1321 ATGAATGAGAAAGAGCATCTTATGTCGACCTGAAATGGAAAGGCTATTGTTTCCCA 1380
Oy      649 ValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArg 668
Db      1381 GTCCACTTACTGACGTCTCCAAACATGAGCCGCTGGCGAGGACAGACGCTGTGGCGCG 1440
Oy      669 SerTPSerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPheAlaCys 688
Db      1441 TCATGAGACGTGCCAGTATCCGCCACTCTTGGCTCGCTGAAGAGATTTTTCGCTGT 1500
Oy      689 Val 689
Oy      1501 GTG 1503

```

```

RESULT 13
US-09-720-086-2
; Sequence 2, Application US/09720086
; Publication No. US20060084053A1
; GENERAL INFORMATION:
; APPLICANT: Li, En
; APPLICANT: Okano, Masaki
; APPLICANT: Xie, Shaoping
; TITLE OF INVENTION: De Novo DNA Cytosine Methyltransferase Genes,
; FILE REFERENCE: 0609,4560002
; CURRENT APPLICATION NUMBER: US/09/720,086
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/14373
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/090,906
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/093,993
; PRIOR FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4195
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-720-086-2

```

```

Alignment Scores:
Pred. No.: 4,92e-232 Length: 4195
Score: 2169.50 Matches: 404
Percent Similarity: 71.9% Conservative: 100
Best Local Similarity: 57.6% Mismatches: 156
Query Match: 57.7% Indels: 41
DB: 5 Gaps: 6

```

```

US-10-623-813-86 (1-689) x US-09-720-086-2 (1-4195)
Oy      10 ProGlyLeuSerGlnLysValGluValAlaSerProProAlaValGlnGlnProThrAsp 29
Db      797 CTTGCCAGGCTCGACTTCATGAGAAAGTAGTACATTAAGGCGGTACATCCCA----- 850
Oy      30 ProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLys 49
Db      851 -----TCAGTTGACTTGAGCCAGCAGATGAGATGAGAGGGTATGATACACACAGGTG 904
Oy      50 AsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIle 69
Db      905 GATCGAAGACAGACAGATGAGTACAGACACAGACAGATATCAGATATTAAGAGTTGGATA 964
Oy      70 GlyGluLeuValTTPGlyLysLeuArgGlyPheSerThrTPProGlyArgIleValSer 89
Db      965 GGTACCTCGTGGGGAAGATCAAGGGCTTCTCCGTGGTGGCTGCCATGTGTGTTC 1024
Oy      90 TTPTPMetThrGlyArgSerArgAlaAlaGluGlyThrArgTPValMetTTPPheGly 109
Db      1025 TGGAAACCCACCTCCACAGCAGACAGCCATGCCGGAATGGCGTGGTACAGTGTGGT 1084
Oy      110 AspGlyLysPheSerValValCysValGluLysLeuMetProLeuSerSerPheCysSer 129
Db      1085 GATGCAAGTTTCTGAGATCTGCTGACAAACTGCTGGCTGGGCTGTTCAAGCCAG 1144
Oy      130 AlaHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluVal 149
Db      1145 CACTTATCTGCTTACTTCATTAACCTGCTTTCTTATGAAAGGCGCATGTACACACT 1204
Oy      150 LeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCysHisAspSerAspGlu 169
Db      1205 CTGAGAAAGCCAGGCTTCAGAGCTGGCAAGACCTC-----TCC 1243
Oy      170 SerAspThrAlaLysAlaValGluValGlnAsnLysProMetIleGluTTPAlaLeuGly 189
Db      1244 AGCAGTCTCGAGAGTCACTGAGAGACCAAGCTGAAGCCCATGTGAGTGGGCCACAGGT 1303
Oy      190 GlyPheGlnProSerGlyProLysGlyLeuGluPro----- 201
Db      1304 GGCCTTCAAGCTTACTGGAGATCGAGGGCTCAAAACCAAGAAAGCAACCACTGTTAAT 1353
Oy      202 -----ProGluGluGlnLysAsnPro 208
Db      1364 AAGTCGAAGTGCCTGCTCAGACAGTAGGAACCTTAGACCCAGGAGGAGAACAA 1423
Oy      209 TyrLysGluValTyrThrAspMetTTPValGluProGluAlaAlaIleTyrAlaProPro 228
Db      1424 AGTCGAAGACGACCAACCAAT-----GACTGCTGCTTGTGACTCCCC 1468
Oy      229 ProProAlaLysLysProArgLysSerThrAlaGluLysProLysValLysGluIle 248
Db      1469 CCACCCCAAGGCGCTCAAGAACATAGCTATAGCGCGGAAGACCGAGGGAG----- 1519
Oy      249 AspGluArgThrArgGluArgGluValTyrGluValArgGlnLysCysArgAsnIleGlu 268
Db      1520 GATAGAGAGACCGAAGACGAGATGCGCTTCGAAGTCAACAACAAGGGAATCTGGAA 1579
Oy      269 AspIleCysIleSerCysGlySerLeuAsnValThrLeuGluHisPProLeuPheValGly 288
Db      1580 GACCGCTGTTGCTCTGTGAAAGAAAGAACCTGTGCTTCCACCCCTCTTTGAGGGGT 1639
Oy      289 GlyMetCysGlnIleAsnLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAsp 308
Db      1640 GGGCTGTGAGAGTTCGCGGATCGCTTCTGAGAGCTCTTTCATGATGATGAGAGAC 1699
Oy      309 GlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsn 328
Db      1700 GGCATATCAGTCTACTACCTCACCGTGTGCTGAGGGCGGTGAACGTGCTGTGAGTAAAC 1759
Oy      329 AsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyVala 348
Db      1760 ACAAGCTGTGCAAGATGCTCTGTGTGTGAGTGTCTGAGAGGTGTGTGGGCGCAGGACA 1819

```





```

US-10-623-813-86 (1-689) x US-09-720-086-4 (1-4145)
QY 13 SerGlnLysValGluGlnLysLeuProPheProAlaValGlnProThraspProAlaSer 32
Db 632 ACACAGAGGACACACACTGGGACGCCCAAGACGACGTACCCCTTACG---CCGCGCTTACG 688
QY 33 ProThrValAlaThr-----ThrProGluProValGlySerAspAlaGlyAsp 48
Db 689 CCCAGAGACGCGACGAGGGGGGATGAGATCCCGGAG--GTGAGAGCAGACAGTGGAGAT 747
QY 49 LysAsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGly 68
Db 748 -----GGAGACAGTTCAGAGTATCAGAGTGGAGAGGATTGGGA 786
QY 69 IleGlyGluLeuValTyrGlyLysLeuAspGlyPheSerTyrProGluArgIleVal 88
Db 787 ATAGGGAGACTGTGTGGGAAAGATCAAGGGCTTCTCGTGGGCGCCCACTGGGGTGG 846
QY 89 SerTyrProMetThrGlyArgSerArgAlaAlaGluGlyThrArgTyrValMetTyrPhe 108
Db 847 TCTTGGAGGCCACCTCCAGAGGACGAGCTATGTCTGGCATCGGGTGGTCCAGTGGTTT 906
QY 109 GlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeuSerSerPheCys 128
Db 907 GCGCATGGCCAAAGTCTCCGAGGCTCTCCGACAAACTGTGGCACTGGGCGCTGTACAC 966
QY 129 SerAlaPheHisGluAlaThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGlu 148
Db 967 CAGCATTTAATTGTGGCCACTTCATTAAGCTGTCTCTATCGAAACCCATGTACAT 1026
QY 149 ValIleGlnValAlaSerSerArgAlaGlyLysLeuPheProValCysHisAspSerAsp 168
Db 1027 GCTCTGGAGAAAGCTTAGGGTGGAGCTGGCAAGACTTCC----- 1068
QY 169 GluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIleGluTyrAlaLeu 188
Db 1069 ---AGCACCCCTGGAGACTCATTTGAGAGCACAGCTGAAGCCCATGTGTGAGTGGGCCAC 1125
QY 189 GlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGluGlu----- 204
Db 1126 GGGGGCTTCAAGCCACTGGGATCGAGGGCTCAAAACCCAAACACGCAACCGAACCTAGTGT 1185
QY 205 -----GluLysAsnProTyrLysGlu 211
Db 1186 AATAAGTCGAAGCTCGCTCGTCAGGACGATGAAGAAATTAGATCAAGAAATACGAGAAC 1245
QY 212 ValTyrThrAspMetTyrValGluProGluAlaAlaAla---TyrAlaProProPro 230
Db 1246 AAGACTCGAAGACGACAGCTGACGACTCAGCCACCTCTGACTGCTGCCCCCGCACCAAG 1305
QY 231 AlaLysLysProArgLysSerThrAlaGluLysProLysValLysGluIleIleAspGlu 250
Db 1306 CGCCTCAAGACAAATTGCTATTAAACAAGCGCAAAAGCCGAGGGGATGA-----GAT 1356
QY 251 ArgThrArgGluArgLeuValTyrGluValAlaArgGlnLysCysArgAsnIleGluAspIle 270
Db 1357 CAGAGCCGAGAAACAATGCTTCAGATGTTCGCCAACACAAAGACGACCTGGAAGATGGC 1416
QY 271 CysIleSerCysGlySerLeuAsnValThrIleGlnHisProLeuPheValGlyGlyMet 290
Db 1417 TGTGTGCTCTTGCGGAGGAAAAAACCCTGCTCTTCAACCTCTCTTGAAGGGGGGGCTC 1476
QY 291 CysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspAspGlyTyr 310
Db 1477 TGTACAGACATCCGGGATCGCTTCTTGAAGCTGTTTACATGATATGATGACATGGCTAT 1536
QY 311 GlnSerTyrCysThrIleCysCysGlyArgGluValLeuMetCysGlyLysAsnAsn 330
Db 1537 CAGCTTACTGCACTGTGTGCTGCGAGAGGCCGAGAGCTGCTCTTTCAGACAAACAGAC 1596
QY 331 CysCysArgCysPheCysValGluCysValAlaPheLeuLeuValGlyProGlyAlaIleGln 350
Db 1597 TGTCTCCGCTGTCTTCTGTGTGAGTGCCTCGAGAGTGTCTGTGGACAGGACAGCGGCC 1656

```

```

QY 351 AlaAlaIleLysGluAspProTyrAsnCysTyrMetCysGlyLysIleLysGlyThrTyrGly 370
Db 1657 GAGGCCAAGCTTCAGAGACCCCTGGAGCTGCTACATGTGTCTCCCGACAGGCTGATGAC 1716
QY 371 LeuLeuArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsnHis 390
Db 1717 GTCTGCGGCGCGGAAAGAGATGAAACGTGGCCCTCGACAGGCTCTTCTTCAACGACAG 1776
QY 391 AspGlnGluPheAspProProLysValTyrProProValProAlaGluLysArgLysPro 410
Db 1777 GGGCTTGAATACGAGGCCCCCAAGCTGACCTGCTGCAATCCCGCAGCCCGAAAGGGCGCC 1836
QY 411 IleArgValIleSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeu 430
Db 1837 ATTCAGAGTCTCTCATTTGTTGATGAGCATCGGACAGGCTACTGCTCTCAAAAGATTG 1896
QY 431 GlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThrValGly 450
Db 1897 GGCATTAAGGTAGAAAGTACGTCGCTTCTGAAGTGTGAGAGATCCATTCGTGTGGA 1956
QY 451 MetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGlnLys 470
Db 1957 ACCGTGAAGCACAGAGGGAGATATCAATACGTGAACGACGTGAGAAACATTCACAAAGAA 2016
QY 471 HisIleGlnGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 490
Db 2017 AATATTGAAGATGGGGCCCATTTGATCTTGATGTGGCGGAAGCCATGACATGACATTC 2076
QY 491 SerIleValAsnProAlaArgLysGlyLeuTyrGlnGlyThrGlyArgLeuPheGlu 510
Db 2077 TCAAAATGTGAATCCAGCCAGAGAAAGCTGTATGAGGGTACAGGCCGCTCTTCTTGGAA 2136
QY 511 PheTyrArgLeuLeuHisAspAlaArgProLysGluGluGlyAspAspArgProPheThr 530
Db 2137 TTTTACCACTCGCTGAATTACTCACGCCCCCAAGAGGGGTGATGACGGCGCTTCTTCTGG 2196
QY 531 LeuPheGluAsnValAlaIleMetGlyValSerAspLysArgAspIleSerArgPheLeu 550
Db 2197 ATGTTTGAGAAATGTTGTAGCCATGAGAGTTGGCCCAAGAGGACATCTCACGGTTCG 2256
QY 551 GluSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHisArgAlaArgTyr 570
Db 2257 GAGTGTAAATCCAGAGATGATGATGATGCCATCAAGTTTCTGCTGCTCACAGGCCCGCATAC 2316
QY 571 PheTyrGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeu 590
Db 2317 TTCTGGGGCAACCTAACCGGAGATGAACAGGCCCGTGAATGACATCAAGAGATGATAAATC 2376
QY 591 GluLeuGlnGluCysLeuGlnHisGlyArgIleAlaLysPheSerLysValArgThrIle 610
Db 2377 GAGCTGCAAGACTGCTTGGATACATAGATAGCAAGTTAAAGAAATACAGACAAATA 2436
QY 611 ThrThrArgSerAsnSerIleLysGlnLysAspGlnHisPheProValPheMetAsn 630
Db 2437 ACCACCAAGTGAACCTCATGATCAACAGGGGAAACCAACTTTTCCCTGTTGTCTGTAAT 2496
QY 631 GluLysGluAspIleLeuTyrCysThrGluMetGluArgValPheGlyPheProValHis 650
Db 2497 GGCAGAAAGATGTTGTGTGTGACATGAGCTCGAAAGATCTTGGCTTCTCTGTGCAC 2556
QY 651 TyrThrAspValIleAsnMetSerArgLeuAlaArgGlnArgLeuGlyArgSerTyr 670
Db 2557 TACACAGAGCTGTCCACATGAGCGGTGTGTGTCGCCCGCAAGACTGCTGGAAAGTCTCTGG 2616
QY 671 SerValProValIleArgHisLeuPheAlaProLeuLysGlyTyrPheAlaCys 688
Db 2617 AGCGGCTGTCAATCGACACTTTCGCCCTTGAAGACTACTTGTGATGT 2670

```

Search completed: October 6, 2006, 04:08:05

Job time : 2168 secs